Longitudinal brain changes in MDD during emotional encoding

Ai, Hui; Opmeer, Esther M; Marsman, Jan-Bernard C; Veltman, Dick J; van der Wee, Nic J A; Aleman, André; van Tol, Marie-José

Published in:
Psychological Medicine

DOI:
10.1017/S0033291719001259

Publication date:
2020

Psychological Medicine

Longitudinal brain changes in MDD during emotional encoding: effects of presence and persistence of symptomatology

--Manuscript Draft--

Manuscript Number:	PSM-D-18-00724R1
Full Title:	Longitudinal brain changes in MDD during emotional encoding: effects of presence and persistence of symptomatology
Article Type:	Original Article
Corresponding Author:	Hui Ai University Medical Center Groningen Groningen, Groningen NETHERLANDS
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University Medical Center Groningen
Corresponding Author's Secondary Institution:	
First Author:	Hui Ai
First Author Secondary Information:	
Order of Authors:	Hui Ai
	Esther M Opmeer
	Jan-Bernard C. Marsman
	Dick J. Veltman
	Nic J.A. van der Wee
	André Aleman
	Marie-José van Tol
Order of Authors Secondary Information:	
Manuscript Region of Origin:	NETHERLANDS
Abstract:	Background: The importance of the hippocampus and amygdala for disrupted emotional memory formation in depression is well-recognized, but it remains unclear whether functional abnormalities are state-dependent and whether they are affected by persistence of depressive symptoms. Methods: Thirty-nine patients with major depressive disorder (MDD) and twenty-eight healthy controls (HC) were included from the longitudinal functional magnetic resonance imaging (fMRI) sub-study of the Netherlands Study of Depression and Anxiety. Participants performed an emotional word-encoding and -recognition task during fMRI at baseline and two-year follow-up measurement. At baseline, all patients were in a depressed state. We investigated state-dependency by relating changes in brain activation over time to changes in symptom severity. Furthermore, effect of time spent with depressive symptoms in the two-year interval was investigated. Results: Symptom change was linearly associated with higher activation over time of the left anterior hippocampus extending to the amygdala during positive and negative word-encoding. Especially during positive word encoding, this effect was driven by symptomatic improvement. There was no effect of time spent with depression in the two-year interval on change in brain activation. Results were independent of medication- and psychotherapy-use. Conclusion: Using a longitudinal within-subjects design we showed that hippocampal-amygdalar activation during emotional memory formation is related to depressive symptom severity but not persistence (i.e. time spent with depression or ‘load’), suggesting functional activation patterns in depression are not subject to functional ‘scarring’ although this hypothesis awaits future replication.

Longitudinal brain changes in MDD during emotional encoding: effects of presence and persistence of symptomatology

Hui Ai[1,2*], Esther M. Opmeer[2], Jan-Bernard C. Marsman[2], Dick J. Veltman[3,4], Nic J.A. van der Wee[5,6], André Aleman[1,2], Marie-José van Tol[2*]

[1.] Shenzhen Key Laboratory of Affective and Social Neuroscience, Shenzhen University, Shenzhen, China.

[2.] University of Groningen, University Medical Center Groningen, Department of Neuroscience, Groningen, The Netherlands.

[3.] VU University Medical Center Amsterdam, Department of Psychiatry, The Netherlands.

[4.] VU University, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands.

[5.] Leiden University Medical Center, Department of Psychiatry, The Netherlands.

[6.] Leiden University, Leiden Institute for Brain and Cognition, Leiden, The Netherlands.

Word count for the text (includes Introduction, Methods, Results and Discussion): **4592**

Word count for abstract: **234**

Number of figures: 1

Number of tables: 2

[*] Corresponding authors: Marie-José van Tol, PhD, Hui Ai, PhD

Neuroimaging Center, University Medical Center Groningen

Antonius Deusinglaan 2, 9713AW Groningen, The Netherlands

Telephone: +31(0) 50 363 8792; Fax: +31(0)50 363 8875

E-mail: m.j.van.tol@umcg.nl, hui.ai@hotmail.com

Abstract

Background: The importance of the hippocampus and amygdala for disrupted emotional memory formation in depression is well-recognized, but it remains unclear whether functional abnormalities are state-dependent and whether they are affected by persistence of depressive symptoms.

Methods: Thirty-nine patients with major depressive disorder (MDD) and twenty-eight healthy controls (HC) were included from the longitudinal functional magnetic resonance imaging (fMRI) sub-study of the Netherlands Study of Depression and Anxiety. Participants performed an emotional word-encoding and -recognition task during fMRI at baseline and two-year follow-up measurement. At baseline, all patients were in a depressed state. We investigated state-dependency by relating changes in brain activation over time to changes in symptom severity. Furthermore, effect of time spent with depressive symptoms in the two-year interval was investigated.

Results: Symptom change was linearly associated with higher activation over time of the left anterior hippocampus extending to the amygdala during positive and negative word-encoding. Especially during positive word encoding, this effect was driven by symptomatic improvement. There was no effect of time spent with depression in the two-year interval on change in brain activation. Results were independent of medication- and psychotherapy-use.

Conclusion: Using a longitudinal within-subjects design we showed that hippocampal-amygdalar activation during emotional memory formation is related to depressive symptom severity **but not persistence (i.e. time spent with depression or 'load')**, suggesting **functional activation patterns in depression are** not subject to functional 'scarring' **although this hypothesis awaits future replication**.

Introduction

Major depressive disorder (MDD) is a prevalent psychiatric disorder associated with high morbidity and mortality, frequently characterized by a chronic or recurrent course (Kessler *et al* 2005). Biased emotional memory has been proposed as a key factor for the development and maintenance of MDD (Ai *et al* 2015, Disner *et al* 2011, Everaert *et al* 2015, Leppänen 2006) and may even underlie the vulnerability for depressive psychopathology (Chan *et al* 2007). Cross-sectional studies suggested that emotional memory biases are state-independent phenomena: better memory for negative information and worse memory for positive information have been reported during both the acute depressive state and during remission (reviewed elsewhere (Bradley and Mathews 1988, Elliott *et al* 2010)), mirroring functional brain abnormalities observed in areas critical for memory formation of emotional material, i.e. the amygdala and hippocampus (Arnold *et al* 2011, Ramel *et al* 2007, van Tol *et al* 2012). Previously, we however observed hyperactivation of the anterior hippocampus/amygdala during encoding of negative information in acutely depressed patients but not in remitted patients in a cross-sectional comparison (van Tol *et al* 2012), suggesting state-dependency instead. However, cross-sectional studies do not allow strong inferences on state-dependency. Importantly, identifying state-dependent neurocognitive markers of MDD may constitute a first step in understanding mechanisms of recovery versus maintenance of depression (Dohm *et al* 2017, Maalouf *et al* 2012, Mayberg 1997).

While longitudinal neuropsychological studies have found that memory biases resolve upon recovery after treatment (Calev *et al* 1986, Peselow *et al* 1991) (though not consistently in Sternberg and Jarvik 1976), functional neuroimaging studies reported mostly changes in activation of the amygdala and hippocampus following symptomatic improvement during affective processing (i.e., not in the context of memory processing) or rest. Findings have been inconclusive with reports of decreased (Fu *et al* 2004, Sheline *et al* 2001, **Redlich *et al* 2017**), increased (Goldapple *et al* 2004, Neumeister *et al* 2006, Ritchey *et al* 2011, Victor *et al* 2010), or unchanged (Fu *et al* 2015, Opmeer *et al* 2015) activation following successful short-term pharmacological treatment (Fu *et al* 2004, Sheline *et al* 2001, Victor *et al* 2010,·Fu *et al* 2015), **electroconvulsive therapy (Redlich *et al* 2017),** cognitive behavioral treatment (Fu *et al* 2008, Goldapple *et al* 2004, Ritchey *et al* 2011), or naturalistic remission (Opmeer *et al* 2015). Heterogeneity in findings may be partly explained by methodological factors such as small sample size, type of stimuli, effects of the (pharmacological) treatment itself on blood flow, or clinical variation in terms of comorbidity or interval between pre- and post-measurement. Nevertheless, the effects of symptomatic improvement on the neural underpinnings of emotional memory processing have not been studied to date.

Because duration of depression has been associated with more severe structural abnormalities, especially in the hippocampus (Frodl *et al* 2008, MacQueen *et al* 2003, Schmaal *et al* 2015), persistence of depressive symptoms may be an important additional factor that influences activation of brain areas important for encoding of emotional information. Such 'persistence' effects may be related to glucocorticoid-dependent toxic effects of stress (Fossati *et al* 2004) and may result in explicit memory deficits (Sapolsky 2000). On a functional level, medial prefrontal involvement during processing of autobiographical memory was found to be blunted in remitted MDD patients but not in individuals at high-risk for developing MDD (Young *et al* 2015), suggesting that memory deficits may be a consequence of having experienced a depressive episode. However, to our knowledge, it has not yet been investigated whether persistence of symptoms may modulate longitudinal functional brain changes related to memory formation.

In the present longitudinal imaging study, we aimed to investigate whether changes in activation of the amygdala and hippocampus during emotional memory encoding are dependent on changes in depressive state and time spent with depressive symptoms. Healthy and depressed participants underwent functional magnetic resonance imaging (fMRI) twice in the context of

the naturalistic and observational Netherlands Study of Depression and Anxiety (NESDA) study, with approximately two years in between. In this interval, **no specific treatment was delivered as part of the study protocol. Given the naturalistic design of our study, participants could receive treatment as usual, which was reconstructed retrospectively based on self-reports at the two-year follow-up interview.** We hypothesized that changes in activation in the hippocampus/amygdala are 1) associated with change in depressive state, especially during negative word encoding and 2) affected by time spent with depressive symptoms between measurements. **Furthermore, we aimed to** explore whether activation in regions **other than amygdala and hippocampus** related to longitudinal treatment responses was associated with severity and time spent with depressive symptoms.

Methods and materials

Participants

Participants were recruited from the ongoing neuroimaging sub-study of the Netherlands Study of Depression and Anxiety (NESDA)(Penninx *et al* 2008) and underwent fMRI scanning at the University Medical Center Groningen (UMCG), Academic Medical Center (AMC) of the University of Amsterdam, and the Leiden University Medical Center (LUMC). NESDA has been designed as a longitudinal observational cohort study with measurements at baseline, one-, two-, four-, six-, and nine-year follow-up, with MRI-measurements performed in a subsample at baseline, two- and nine-year follow up (nine-year follow-up measurement was completed during the preparation of this manuscript). At baseline, patients with MDD (n=70), MDD and one or more anxiety disorders (i.e. social anxiety disorder (SAD), panic disorder (PD) and/or generalized anxiety disorder (GAD); N=92), patients with only anxiety disorders (i.e. SAD, PD, and/or GAD; n=71), and healthy control participants (HC; n=68) were included. The ethical review board of each participating center approved the study and all participants gave written informed consent.

Exclusion criteria for all participants in the NESDA neuroimaging study at baseline (n=301) were: age under 18 or over 57 years; current alcohol or substance abuse; presence or history of a neurological or somatic disorder with possible effects on the central nervous system; general 3T MRI contraindications; hypertension. Use of selective serotonin reuptake inhibitors (SSRIs) or infrequent use of benzodiazepines (oxazepam [max 20 mg] or diazepam, maximum of three times a week and not within 48 hours before scanning) was allowed. Patients using any other psychopharmacological agent were excluded. Exclusion criteria for the second measurement at two-year follow-up (S2; N=199) were identical, with the exception of the age criterion. Also, from a cohort perspective, we were less strict in excluding patients based on type of medication used at S2 (see Table 1 and Supplementary Table S1 for details). In line with the observational nature of the NESDA study, no specific treatment was delivered in between measurements, but was monitored retrospectively. Participants were free to consult their general practitioner, psychiatrist or psychologist for the help they wished to receive. Results of the baseline measurement (S1) and their associations with subsequent course related to emotional memory processing have been published elsewhere (Ai *et al* 2015, van Tol *et al* 2012).

Complete behavioral data and good quality fMRI data at both S1 and S2 were available of 64 MDD patients and 39 HC. At S1, all patients fulfilled the criteria for a diagnosis of MDD with a half-year recency based on the Composite International Diagnostic Interview (CIDI life time - version 2). An additional diagnosis of SAD, PD and/or GAD at either S1 or S2 was allowed (See Table 1 for details). Following Opmeer *et al.*, (Opmeer *et al* 2015), we included only patients who were in a depressive state at S1 defined as a Montgomery–Åsberg Depression Rating Scale (MADRS) score larger than 10 (Zimmerman *et al* 2004). One participant had a huge increase in MADRS score at S2 and was classified as an outlier (change score >3SD from group mean) and subsequently excluded from the analyses. The final patient sample included 39 individuals. In total, 11 HC were excluded from further analysis based on the presence of possible depressive symptomatology at S2 (i.e. MADRS-score >10; n=1), too high level of education to be matched to the patient group (n=1) or unreliable task performance (n=9; Supplementary Figure S1). This resulted in the inclusion of 28 HC without any current or life-time DSM-IV diagnosis and no indication of depressive symptomatology at both S1 and S2 (See Supplementary Figure S1 for a flow diagram reflecting data selection).

Task paradigm

All participants performed the event-related, subject-paced, emotional word encoding and recognition task during both fMRI scanning sessions (S1 and S2)(van Tol *et al* 2012). During the encoding phase, 20 blocks containing 160 stimuli (positive/ neutral/ negative words and baseline trials; 40 each) were pseudo-randomly presented. Participants were instructed to evaluate whether the word was positive, negative or neutral in valence by pressing the right, left and middle button, respectively. During baseline trials, participants were asked to press the corresponding button to indicate the direction of the arrow. After a retention interval of 10 minutes (during which the structural T1 scan was acquired), the retrieval phase started and consisted of 120 encoding target words, 120 distracter words and 40 baseline words that were presented in 20 pseudo-randomized blocks. Participants were instructed to indicate whether they had seen, had not seen, or probably had seen the word. Emotional words in the valence categories were matched based on length, frequency in the Dutch language and complexity. The same words list was used in both measurements, although the order was changed at the two-year follow-up measurement. The emotional word encoding task was preceded by an executive planning task (van Tol *et al* 2011) and followed by an emotional face viewing task (Demenescu *et al* 2011, Opmeer *et al* 2015) and a resting state acquisition (Veer *et al* 2010). Based on the hypotheses formulated in our cross-sectional study (van Tol *et al* 2012), we only investigated the encoding session.

fMRI data acquisition

Neuroimaging data were collected with 3T Philips MR-scanners located in Amsterdam, Leiden, and Groningen using standard EPI techniques, though with minor differences in acquisition parameters. A detailed description of acquisition specifications can be found in the supplemental material.

Data analysis

Independent variables

Firstly, to test for the correlation between symptom change and brain activation change over time, a relative symptom change score representing the difference in depression severity between S1 and S2 while taking into account baseline severity was calculated for each patient (i.e., [MADRS S2 – MADRS S1]/MADRS S1). Furthermore, to be able to compare changes over time in behavior and brain activation following symptomatic change with changes in HC, who were also scanned twice, and to explore e.g. whether change in the high improved patients represented normalization (i.e., approached activation of HC at S2) or whether change in low improved patients represented further deviations from normal, we divided the patients in two groups based on the median of relative symptom change scores (median = -.46): a group of high improved (MDD-HI; n=20, Supplementary Figure S1) and a group of low improved patients (MDD-LI; n=19).

Secondly, to test for the correlation between brain activation change and percentage of time spent with depression (i.e. persistence), presence of depressive symptoms per month for the duration of the interval between S1 and S2 was assessed with the life chart interview (Lyketsos *et al* 1994) at S2. Participants had to rate the severity of depressive symptoms per month and only symptoms with small to severe burden were taken as indication of presence of symptoms. Percentage of months experiencing depressive symptoms **relative to the overall follow-up period** was calculated per patient as time spent with depression (Ai *et al* 2015).

Clinical variables and behavioral data

Effects of symptom change and time spent with depressive symptoms on demographic, psychometric assessment and memory performance were analyzed in IBM SPSS software (SPSS v.22.0, IBM). We employed analyses of covariance (ANCOVA), Chi-square tests and *t*-tests where appropriate for demographic and psychometric data with a significance level of $p<.05$, two-tailed.

For the behavioral data, performance difference scores (S2-S1) for both reaction times (RT) and accuracy for successfully encoded words (Tulving 1985) were calculated. We assessed the continuous association between relative symptom change scores and depressive duration, and RT and accuracy difference scores over time in patients. Age and years of education were included as covariates. A sensitivity analysis was performed within patients who showed symptomatic improvement (thus, patients who were equally or more depressed at S2 than at S1 were excluded; n=6).

Additionally, to investigate whether patients (MDD-HI/LI) performed differently over time as compared to HC, we set up a group (3; HC, HI, LI) × valence (3; positive, negative, neutral) × time (2; S1, S2) repeated measures ANCOVA, with age and years of education as covariates. Effects were considered significance at $p<.05$. Where appropriate, Bonferroni correction for multiple comparisons was applied.

Imaging data preprocessing

For the fMRI data, preprocessing and task modeling was performed with Statistical Parametric Mapping software (SPM8, Wellcome Trust Center for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm) implemented in Matlab 7.8 (The Math Works Inc., Natick, MA, USA). A detailed description of the preprocessing steps and first-level modeling can be found in the supplemental material.

Effects of change of depressive state

To test for the association between symptom change and change of brain activation during positive and negative encoding over the two-year interval, scan moments, S2-S1 contrast maps were entered as dependent variables in a full-factorial model, with valence (**successfully encoded positive words>successfully encoded neutral words** [S2-S1], **successfully encoded negative words>successfully encoded neutral words** [S2-S1]) as interacting factor with valence. **Contrast maps were built for successful encoding of positive words (vs. successful encoding of neutral words) and negative words separately.** To control for the possible confounding effects of variations within and between participants in scanning site (which coincided with minor variations in sequence and coil; see supplementary material), four dummy variables for site (i.e., both times scanned in AMC; changed from AMC to LUMC; changed from LUMC to AMC; both times scanned in UMCG; both times scanned in LUMC) were defined as covariates of no interest. In addition, age and years of education at S1 were added as covariates.

We repeated our analysis with the following possible confounding factors added separately to the model: percentage of time spent with depression, relative changes in anxiety severity assessed by Beck Anxiety Inventory ([BAI-scores S2-S1]/ BAI-scores S1) (Beck *et al* 1988), SSRI-use, and participation in psychotherapy. SSRI-use at/between S1 and S2 was added to the model by means of three dummy variables (used at both S1 and S2, started after S1, stopped after S1, both not used). Psychotherapy-use between S1 and S2 was coded as a dummy variable and added as covariate to test for the effect of psychotherapy. Use of SSRI and psychotherapy between S1 and S2 are summarized in Supplementary Table S1.

A sensitivity analysis was planned to test whether associations **would hold in the analysis including only** patients with symptomatic improvement **(n=33)**.

Effects of persistence of depressive symptoms

We built a full factorial model with valence as factor (2; **successfully encoded positive words>successfully encoded neutral words and successfully encoded negative words>successfully encoded neutral words**) and time spent with depressive symptoms as an interacting covariate with valence. Site (four dummy variables), age, and years of education were added as covariates. We tested for the effects of time spent with depressive symptoms during encoding of positive words and negative words separately. In a subsequent step, relative symptom change of depressive and anxiety symptom severity and treatment at S1 and S2 (medication-and psychotherapy use; yes/no) were added separately as covariates to statistically control for their possible confounding effects.

A sensitivity analysis was planned within patients with symptomatic improvement.

Statistical thresholding

Based on previous studies (see introduction), we *a priori* defined the bilateral hippocampus and amygdala as our regions-of-interest (ROI) and built one composite mask encompassing these regions. The regions were defined according to the automated anatomical labels of the Wake Forest University (WFU, Winston Salem, North Carolina) Pick Atlas toolbox. Small volume correction for multiple comparisons was applied within the ROI. F-tests in the main and follow-up analyses were explored separately for positive and negative words at $p<.001$ uncorrected. *Post hoc* t-tests were regarded significant at a threshold of $p<.05$ family wise error (FWE) corrected at voxel-level (with an initial threshold of $p<.001$ uncorrected). **We also examined the effects in other brain regions than ROIs, which** had to meet $p<.05$, FWE whole**-**brain correct**ed** to be considered significant.

Results

Demographic characteristics

Demographics and clinical characteristics of all patients and healthy controls are summarized in Table 1 and supplementary material. Clinical characteristics of high-improved and low-improved patient groups that were included in explorative post-hoc analyses are listed in Supplementary Table and Supplementary results. Thirty-three patients showed symptomatic improvement (S2<S1), two remained stable (S2=S1) and four showed more severe symptoms at S2 (S2>S1).

Behavioral results

No correlations were found between relative depressive symptom change and changes in performance on memory of positive, neutral or negative words over time (i.e., RTs and accuracy) (p>.05). Sensitivity analyses within symptomatically improved patients only (n=33) did not change this result. Group × Time repeated measures ANOVA indicated no changes neither in performance and response times in HC nor a difference between HC and HI or LI (ps>.05).

There was no association between time spent with depressive symptoms and changes in behavioral performance (p>.05).

fMRI results

Correlations with change of depressive state

Relative symptom change was negatively correlated with activation change in the bilateral hippocampal/amygdala during both positive and negative word encoding (Table 2; Figure 1). However, only the effect in the left hippocampus survived multiple comparison correction and indicated that larger symptomatic improvement coincided with a larger increase in left anterior hippocampal activation during encoding of emotional information.

Adding time spent with depressive symptoms in the interval between S1 and S2 as covariate did not change the results (Z=3.85, p_{FWE}=.019 for **successfully encoded positive words>successfully encoded neutral words** (pos); Z=3.95, p_{FWE}=.014 for **successfully encoded negative words>successfully encoded neutral words** (neg)). Also, results were not affected by including change in anxiety severity as a covariate to the model (pos: Z=3.82, p_{FWE} =.021; neg: Z=3.77, p_{FWE} =.025) or by adding SSRI-use at S1 and S2 as covariates (pos: Z=3.68, p_{FWE} =.034; neg: Z=3.49, p_{FWE} =.06). Results bordered statistical significance after adding psychotherapy as a covariate (pos: Z=3.59, p_{FWE} =.05; neg: Z=3.54, p_{FWE} =.05).

When repeating the analysis in the symptomatically improved patients only, the negative correlation between symptom change and brain activation change in the hippocampus was observed subthreshold (MNI coordinates:[x=-18, y=-13, z=-11], Z=3.51, p_{FWE}=.09) during positive encoding, and was not significant during negative encoding (p_{FWE}=.50).

Furthermore, *post hoc* group comparison (detailed in the supplementary methods and results) showed that activation estimates in our main cluster did not change in HC over time, and plots suggested a trend of normalization during positive but not negative word encoding in the high-improved group (Supplementary Figure 3A&3B).

Correlations with time spent with depressive symptoms and course

No correlation between percentage of time with depressive symptoms and changes in brain activation was observed across all MDD patients during **successful encoding on positive and negative words**. Adding change in depressive and anxiety symptoms or medication/therapy use to the model did not change this observation.

Discussion

In this longitudinal study, we examined changes in emotion-related brain activation over time associated with symptomatic improvement and time spent with depressive symptoms in depressed patients. Symptomatic improvement was associated with increased responses in the anterior hippocampus/amygdala during encoding of emotional stimuli over time. Follow-up explorations indicated that increased activation of the hippocampal/amygdala responsiveness occurred in the direction of normalization, especially for the encoding of positive words. The effect was unrelated to changes in anxiety severity, and use of SSRIs, although it became smaller after adding use of psychotherapy as a covariate. No relation was observed between depression duration (i.e. time spent with depressive symptoms) in the two-year follow-up nor were changes in hippocampal and amygdalar activation observed. These results suggest that hippocampal activation during emotional memory formation changes with symptomatic improvement, but is not subject to functional 'scarring' as a result of enduring symptom manifestation. Our results indicate that symptomatic improvement is at least partially associated with normalization of limbic responsiveness to positive material.

Based on previous reports on memory bias-related brain activation abnormalities in depression (Arnold *et al* 2011, Hamilton and Gotlib 2008, Ramel *et al* 2007, van Tol *et al* 2012, Van Wingen *et al* 2010) and our previous cross-sectional observations (van Tol *et al* 2012), we hypothesized state-dependency of activation of the amygdala and hippocampus specific for negative valence information, and thus changes of activation as a function of symptomatic recovery. In line with this hypothesis, hippocampal reactivity during negative encoding correlated with symptomatic change. Moreover, state-dependency was observed during positive encoding. Although similar linear relations with symptomatic improvement were observed for both positive and negative encoding, changes during positive word encoding showed to be a more specific indicator of symptomatic improvement. This was indicated by the stability of effects when excluding the patients that worsened in terms of symptom severity and by the fact that the *post-hoc* plotting of effects indicated an increase of activation in the improved patients only. This increase followed a pattern of normalization (i.e. approaching activation in the HC). During negative encoding, associations were no longer significant when studied in the symptomatic improved patients only. This suggests that state-dependent changes during positive encoding may be a preferred marker of symptomatic improvement. Notwithstanding, although longitudinal studies did not study emotional encoding for both positive and negative information so far, our study supports findings of altered reactivity to positive information (Fu *et al* 2007, Victor *et al* 2010, Wise *et al* 2014), and suggests normalized reactivity to positive-related effects.

The hippocampus has been proposed as a target for both anti-depressant treatment and cognitive behavioral therapy (CBT) (Goldapple *et al* 2004). Treatment studies have confirmed the importance of the hippocampus by consistently reporting normalization of hippocampal activation following pharmacological treatment (Anand *et al* 2007, Arnone *et al* 2012b, Fu *et al* 2004) and CBT (Goldapple *et al* 2004, Ritchey *et al* 2011). In the current study, we studied the neural characteristics related to naturalistic changes in depressive state, which was not attributable to treatment with antidepressant medication. However, most of our sample received at least one type of psychological care. Therefore, we cannot fully rule out of the effect of psychotherapy and indeed our effects were slightly attenuated when treatment with psychotherapy was added to the model. Together, our observations suggest that increased hippocampal responsiveness to emotional material may not only reflect treatment effects of or symptomatic improvement following anti-depressants or psychological treatment(Fu *et al* 2007, Victor *et al* 2010, Wise *et al* 2014) but also naturalistic improvement.

No other regions were found to change as a function of symptomatic improvement. Although changes in regions such as the ventromedial prefrontal cortex (Ritchey *et al* 2011), anterior cingulate cortex (Fu *et al* 2008, Fu *et al* 2008), frontal pole (Usami *et al* 2014), and the extrastriate cortex (Fu *et al* 2007) have been reported by previous longitudinal treatment studies. They have been reported in the context of emotional processing, but not in the context of memory formation or using verbal stimuli. Additionally, other studies have reported that prefrontal alterations might be a trait marker rather than a state marker of vulnerability to depression (Elliott *et al* 2012, Tomioka *et al* 2015), which was not the focus of our study.

A second aim of this study was to investigate whether time spent with depressive symptoms was associated with greater functional brain alterations during emotional memory encoding. We found that depression duration was not correlated with changes of activation in the hippocampus, which indicates that the neurotoxic or scarring hypothesis might not be relevant to functional changes over time. Previous cross-sectional and longitudinal studies suggested that hippocampal volume is negatively related to duration of illness in MDD, represented by **history of psychiatric hospitalization (Zaremba *et al* 2018)**, number of episodes (MacQueen *et al* 2003, Treadway *et al* 2015) and duration of untreated illness (Sheline *et al* 1999), though not consistently (Bremner *et al* 2000, **McKinnon *et al* 2009**). At the same time, volumetric changes in the hippocampus have been linked to symptomatic improvement following treatment (Arnone *et al* 2012a), suggesting state-dependency of hippocampal volume. In the present study, though patients differed in course trajectory of depression, changes of brain activation were not related to depressive course, indicating that functional longitudinal changes observed in the hippocampus are load-independent. **However, the variety in selected clinical variables of current and previous studies might explain some heterogeneity in reported results.** Together, our results indicate that functional responsiveness of limbic brain regions may be more related to depressive state, without exacerbation of abnormalities as a function of unfavorable course of the depression.

Some limitations of our study should be noted. First, although clear strengths of our study are its longitudinal naturalistic design and that we could control for activation changes over the same interval in a healthy sample, the associations we found between changes of brain activation and symptom change over time are correlational in nature and do not imply causation of remission in depression. **And this effect was not found in a formal group × time × valence interaction. However, testing this was not the aim of our paper because we focused on changes over time within depressed patients.** Second, we investigated symptom severity change of depression rather than symptom remission. Although most of our high-improved patients were recovered at the time of the follow-up measurement, our conclusions cannot be generalized to changes associated with stable remission. Third, although adding SSRI-use and psychotherapy use as covariates to the model did not change the observed relations, this does not fully rule out specific medication/treatment effects. Fourth, caution should be taken in interpreting our result as a true memory effect (i.e., hits-misses), because the number of error trials was too low to investigate this. More sensitive measures on behavioral changes in primary emotional and memory processing are necessary in future studies. **Fifth, although the site effect was controlled by adding it as a covariate, it might still have confounding effect on our results. Quality assurance analysis and exploration by excluding patients that switched scanners between measurements (supplementary results) revealed similar results. These indicate that our observed effects, especially those observed during positive encoding, were not primarily driven by site-specific changes in signal over time. Next, the retrospective life chart method used to measure persistence of depressive symptoms might have been subject to patients' mood state, though the reliability and validity have been estimated to be relatively high (Warshaw, *et al* 2001). Furthermore, although comorbidity of SAD and PD was similar in low and high improved MDD groups, GAD**

was more frequent in low-improved MDD patients, which may have affected our results. Finally, it is possible that the encoding processing was more explicit at S2 than at S1, because people at S2 could have remembered that a recognition phase followed the encoding phase. However, implicit and explicit memory processing have been suggested to be subject to the same encoding factors and rely on similar perceptual processes and representations (Turk-Browne *et al* 2006), which is corroborated by the lack of differences over time in the HC group in our study.

Conclusion

By characterizing longitudinal changes of activation in the anterior hippocampus/amygdala during emotional memory encoding, our study showed that the neural correlates of emotional memory formation change with improvement of the depressive state. Furthermore, our findings suggest a normalization of activation especially for positive information. On the other hand, enduring depressive symptom manifestation was not related to longitudinal changes in hippocampal-amygdalar activation. Taken together, **our results suggest** that hippocampal activation is a state-dependent characteristic that **is not related to persistence of depression. This may indicate that functional activation patterns in depression are** not subject to functional 'scarring', **a hypothesis that deserves further investigation**.

345 **Tables and Figures**

346 **Figure 1.** Brain activation during emotional word encoding. A). Negative association between symptom change and
347 hippocampal activation change during positive word encoding. (peak MNI coordinate: **x=-27, y=-16, z=-11**); B). Negative
348 association between symptom change and hippocampal activation change during negative word encoding. (peak MNI
349 coordinate: **x=-24, y=-13, z=-11**).

350
351

352
353 **Table 1.** Demographics characteristics.

		HC	High-improved MDD	Low-improved MDD	F	t	χ^2	*Likelihood ratio*	*p*
N		28	19	20	-	-	-	-	
Diagnosis over time (Remitted/non-remitted)	N	-	17/2	9/10	-	-	-	-	-
State change over time (improved/stable/worsen)	N	-	19/0/0	14/2/4	-	-	-	-	-
Site S1(AMC/LUMC/ UMCG)	N	15/9/4	8/8/3	8/8/4	-	-	-	1.51	.89
Site S2(AMC/LUMC/ UMCG)	N	13/11/4	7/9/3	8/8/4	-	-	-	.70	.95
Sex (male/female)	N	10/18	7/12	9/11	-	-	.47	-	.79
Age	M(SD)	39.82(9.68)	37.32(9.59)	39.55(11.26)	.38	-	-	-	.68
Years of education	M(SD)	14.46(2.77)	12.37(2.17)	13.60(3.78)	2.83	-	-	-	.07
Months interval	M(SD)	21.85(1.38)	22.63(1.30)	22.20(1.61)	1.66	-	-	-	.20
MADRS_S1	M(SD)	.93(1.44)	19.11(5.17)	21.55(7.33)	127.5	-	-	-	<.001*[1]
MADRS_S2	M(SD)	.50(1.00)	4.16(2.83)	17.90(6.37)	126.0	-	-	-	<.001*[2]
Relative MADRS_S2>S1	M(SD)	-.81(.36)	-.78(.15)	-.14(.29)	-	-8.61	-	-	<.001*[2]
BAI_S1	M(SD)	2.07(2.70)	12.32(7.33)	15.15(9.76)	24.83	-	-	-	<.001*[1]
BAI_S2	M(SD)	2.14(2.03)	7.58(5.61)	14.10(8.50)	26.08	-	-	-	<.001*[2]
Relative BAI_S2>S1	M(SD)	.02(.99)	-.45(.38)	.45(1.92)	-	-1.89	-	-	.07
Depressive duration between S1 and S2 (%)	M(SD)	-	.42(.40)	.58(.40)	-	-1.22	-	-	.23
Months with depressive symptom before S1	M(SD)	-	16.42(14.69)	22.85(16.28)	-	-1.29	-	-	.20
Comorbidity_S1(MDD/MDD$^+$)									
Comorbid SAD	N	-	6/13	9/11	-	-	.74	-	.51
Comorbid PD	N	-	6/13	6/14	-	-	.01	-	.92
Comorbid GAD	N	-	7/12	10/10	-	-	67	-	.52
Comorbidity at follow-up									
Comorbid SAD(yes/no)	**N**	**-**	**2/17**	**6/14**	**-**	**-**	**-**	**2.36**	**.13**
Comorbid PD(yes/no)	**N**	**-**	**2/17**	**6/14**	**-**	**-**	**-**	**2.36**	**.13**
Comorbid GAD(yes/no)	**N**	**-**	**0/19**	**8/12**	**-**	**-**	**-**	**12.66**	**<.01**
Age of depressive onset	M(SD)	-	26.89(11.27)	21.74(9.76)	-	1.51	-	-	.14
# of episodes prior to SI	M(SD)	-	1.36(.67)	1.64(.67)	-	-.95	-	-	.35
Psychotherapy-use_S1	M(SD)	-	4/15	6/14	-	-	-	.41	.71
Psychotherapy-use_S2	M(SD)	-	9/10	6/14	-	-	1.24	-	.33
Psychotherapy-use between S1&S2(both used/stopped after S1/started after S1/both not used)	M(SD)	-	10/5/0/4	12/2/2/4	-	-	-	4.26	.24
SSRI-use_S1(yes/no)	N	-	7/12	7/13	-	-	0.01	-	.91
SSRI-use_S2(yes/no)	N	-	7/12	3/17	-	-	-	2.49	.16
SSRI-Use between S1&S2 (both used/stopped after S1/started after S1/both not used)	N	-	5/2/2/10	2/5/1/12	-	-	-	3.15	.37
Benzodiazepine-use_S2	N	-	4/15	3/17	-	-	-	.24	.62

[1.] HC differed from both patient groups, while the two patient groups did not differ; [2.] All groups differed from each other; [3.] Infrequent use; [4.] Two patients used benzodiazepine frequently. * significant at $p<.05$

HC: healthy control; S-R: symptom-remitted MDD patients; S-S: symptomatic-symptomatic MDD patients; SAD: social anxiety disorder; PD: panic disorder; GAD: generalized anxiety disorder.

354
355

Table 2. Correlation between state-change scores and brain activation changes across patients

	MNI Coordinate									
Regions	k[a]	k[b]	Side	BA	x	y	z	*T*	*Z*	p_{FWE_SVC}
successfully encoded positive words> successfully encoded neutral words*: negative correlation*										
Hippocampus/amygdala	35	13	L	20	**-27**	**-16**	-11	3.83	3.63	.040*
Hippocampus/amygdala	33	9	R	34	27	-4	-11	3.46	3.31	.107
successfully encoded negative words> successfully encoded neutral words: *negative correlation*										
Hippocampus/amygdala	50	22	L	-	-24	-13	-11	3.76	3.57	.049*
Hippocampus/amygdala	59	20	R	-	15	-7	-17	3.40	3.26	.122

[a.] Cluster size in whole-brain analysis; [b.] Cluster size after small volume correction.
* Significant at p<.05 FWE corrected, voxel-level after small volume correction (SVC).

Acknowledgments

This study is supported by the participating universities and mental health-care organizations: VU University Medical Center Amsterdam, University Medical Center Groningen, Leiden University Medical Center, GGZ inGeest, Arkin, GGZ Rivierduinen, Lentis, GGZ Friesland, GGZ Drenthe, Scientific Institute for Quality of Healthcare (IQ healthcare), Netherlands Institute for Health Services Research (NIVEL) and Netherlands Institute of Mental Health and Addiction (Trimbos Institute).

Financial support

Hui Ai is jointly supported by the National Natural Science Foundation of China [grant No.31700959] and Department of Education of Guangdong Province [grant No. 2016KQNCX149] and Department of Science and Technology of Guangdong Province [Guangdong Pearl River Talents Plan for Postdoctoral researchers]. Marie-José van Tol was supported by a VENI grant from the Netherlands Organization for Scientific Research (N.W.O.) [grant no. 016.156.077]. Esther Opmeer is partly supported by grant from N.W.O. ZonMw [grant no. 537001005]. The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development [ZonMw, grant number 10-000-1002] and financial contributions by participating universities and mental health care organizations (VU University Medical Center, GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekscentrum).

Conflict of interest

Hui Ai, Jan-Bernard C. Marsman, Dick J. Veltman and Esther M. Opmeer declare no conflict of interest. Nic van der Wee received speaking fees from Eli Lilly and Wyeth; and served on advisory panels of Eli Lilly, Pfizer, Wyeth and Servier. Marie-José van Tol received speakers fees from Lundbeck n.v. André Aleman received an investigator-initiated unrestricted research grant from Bristol-Myers Squibb and speakers bureau honoraria from Lundbeck n.v. All of these activities are not directly related to the present study and, therefore, do not form a conflict of interest.

Ethical standard

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

387 **References**

388 **Ai H, Opmeer EM, Veltman DJ, van der Wee, Nic JA, van Buchem MA, Aleman A, van Tol M** (2015). Brain Activation
389 During Emotional Memory Processing Associated with Subsequent Course of Depression. *Neuropsychopharmacology*

390 **Anand A, Li Y, Wang Y, Gardner K, Lowe MJ** (2007). Reciprocal effects of antidepressant treatment on activity and
391 connectivity of the mood regulating circuit: an FMRI study. *The Journal of neuropsychiatry and clinical neurosciences* **19**, 274-
392 282

393 **Arnold JF, Fitzgerald DA, Fernández G, Rijpkema M, Rinck M, Eling PA, Becker ES, Speckens A, Tendolkar I** (2011).
394 Rose or black-coloured glasses?: Altered neural processing of positive events during memory formation is a trait marker of
395 depression. *Journal of affective disorders* **131**, 214-223

396 **Arnone D, McKie S, Elliott R, Juhasz G, Thomas E, Downey D, Williams S, Deakin J, Anderson I** (2012a). State-dependent
397 changes in hippocampal grey matter in depression. *Molecular psychiatry*

398 **Arnone D, McKie S, Elliott R, Thomas EJ, Downey D, Juhasz G, Williams SR, Deakin JW, Anderson IM** (2012b).
399 Increased amygdala responses to sad but not fearful faces in major depression: relation to mood state and pharmacological
400 treatment. *American Journal of Psychiatry* **169**, 841-850

401 **Beck AT, Epstein N, Brown G, Steer RA** (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal*
402 *of consulting and clinical psychology* **56**, 893

403 **Bradley BP, Mathews A** (1988). Memory bias in recovered clinical depressives. *Cognition & Emotion* **2**, 235-245

404 **Bremner JD, Narayan M, Anderson ER, Staib LH, Miller HL, Charney DS** (2000). Hippocampal volume reduction in major
405 depression. *American Journal of Psychiatry* **157**, 115-118

406 **Calev A, Korin Y, Shapira B, Kugelmass S, Lerer B** (1986). Verbal and non-verbal recall by depressed and euthymic affective
407 patients. *Psychological medicine* **16**, 789-794

408 **Chan SW, Goodwin GM, Harmer CJ** (2007). Highly neurotic never-depressed students have negative biases in information
409 processing. *Psychological medicine* **37**, 1281-1291

410 **Demenescu L, Renken R, Kortekaas R, van Tol M, Marsman J, van Buchem M, van der Wee N, Veltman D, den Boer J,**
411 **Aleman A** (2011). Neural correlates of perception of emotional facial expressions in out-patients with mild-to-moderate
412 depression and anxiety. A multicenter fMRI study. *Psychological medicine* **41**, 2253

413 **Disner SG, Beevers CG, Haigh EAP, Beck AT** (2011). Neural mechanisms of the cognitive model of depression. *Nature*
414 *Reviews Neuroscience* **12**, 467-477

415 **Dohm K, Redlich R, Zwitserlood P, Dannlowski U** (2017). Trajectories of major depression disorders: A systematic review of
416 longitudinal neuroimaging findings. *The Australian and New Zealand Journal of Psychiatry* **51**, 441-454

417 **Elliott R, Lythe K, Lee R, McKie S, Juhasz G, Thomas EJ, Downey D, Deakin J, Anderson IM** (2012). Reduced Medial
418 Prefrontal Responses to Social Interaction Images in Remitted Depression. *Archives of General Psychiatry* **69**, 37

419 **Elliott R, Zahn R, Deakin JFW, Anderson IM** (2010). Affective cognition and its disruption in mood disorders.
420 *Neuropsychopharmacology* **36**, 153-182

421 **Everaert J, Duyck W, Koster EH** (2015). Emotionally biased cognitive processes: The weakest link predicts prospective
422 changes in depressive symptom severity

423 **Fossati P, Radtchenko A, Boyer P** (2004). Neuroplasticity: from MRI to depressive symptoms. *European*
424 *Neuropsychopharmacology* **14, Supplement 5**, S503-S510

425 **Frodl TS, Koutsouleris N, Bottlender R, Born C, Jäger M, Scupin I, Reiser M, Möller H, Meisenzahl EM** (2008).
426 Depression-Related Variation in Brain Morphology Over 3 YearsEffects of Stress?. *Archives of General Psychiatry* **65**, 1156-
427 1165

428 **Fu CHY, Williams SCR, Cleare AJ, Scott J, Mitterschiffthaler MT, Walsh ND, Donaldson C, Suckling J, Andrew C,**
429 **Steiner H** (2008). Neural responses to sad facial expressions in major depression following cognitive behavioral therapy.
430 *Biological psychiatry* **64**, 505-512

Fu CH, Costafreda SG, Sankar A, Adams TM, Rasenick MM, Liu P, Donati R, Maglanoc LA, Horton P, Marangell LB (2015). Multimodal functional and structural neuroimaging investigation of major depressive disorder following treatment with duloxetine. *BMC psychiatry* **15**, 82

Fu CH, Williams SC, Brammer MJ, Suckling J, Kim J, Cleare AJ, Psych M, Walsh ND, Mitterschiffthaler MT, Andrew CM (2007). Neural responses to happy facial expressions in major depression following antidepressant treatment. *The American Journal of Psychiatry* **164**, 599-607

Fu CH, Williams SC, Cleare AJ, Brammer MJ, Walsh ND, Kim J, Andrew CM, Pich EM, Williams PM, Reed LJ (2004). Attenuation of the neural response to sad faces in major depression by antidepressant treatment: a prospective, event-related functional magnetic resonance imaging study. *Archives of General Psychiatry* **61**, 877

Fu CHY, Williams SCR, Cleare AJ, Scott J, Mitterschiffthaler MT, Walsh ND, Donaldson C, Suckling J, Andrew C, Steiner H, Murray RM (2008). Neural Responses to Sad Facial Expressions in Major Depression Following Cognitive Behavioral Therapy. *Biological psychiatry* **64**, 505-512

Goldapple K, Segal Z, Garson C, Lau M, Bieling P, Kennedy S, Mayberg H (2004). Modulation of cortical-limbic pathways in major depression: treatment-specific effects of cognitive behavior therapy. *Archives of General Psychiatry* **61**, 34-41

Hamilton JP, Gotlib IH (2008). Neural substrates of increased memory sensitivity for negative stimuli in major depression. *Biological psychiatry* **63**, 1155-1162

Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* **62**, 593-602

Leppänen JM (2006). Emotional information processing in mood disorders: a review of behavioral and neuroimaging findings. *Current Opinion in Psychiatry* **19**, 34

Lyketsos CG, Nestadt G, Cwi J, Heithoff K (1994). The Life Chart Interview: a standardized method to describe the course of psychopathology. *International Journal of Methods in Psychiatric Research*

Maalouf FT, Clark L, Tavitian L, Sahakian BJ, Brent D, Phillips ML (2012). Bias to negative emotions: A depression state-dependent marker in adolescent major depressive disorder. *Psychiatry research* **198**, 28-33

MacQueen GM, Campbell S, McEwen BS, Macdonald K, Amano S, Joffe RT, Nahmias C, Young LT (2003). Course of illness, hippocampal function, and hippocampal volume in major depression. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1387-1392

Mayberg HS (1997). Limbic-cortical dysregulation: a proposed model of depression. *The Journal of neuropsychiatry and clinical neurosciences*

McKinnon MC, Yucel K, Nazarov A, MacQueen GM (2009). A meta-analysis examining clinical predictors of hippocampal volume in patients with major depressive disorder. *J Psychiatry Neurosci*. **34**, 41-54.

Neumeister A, Drevets WC, Belfer I, Luckenbaugh DA, Henry S, Bonne O, Herscovitch P, Goldman D, Charney DS (2006). Effects of a α2C-adrenoreceptor gene polymorphism on neural responses to facial expressions in depression. *Neuropsychopharmacology* **31**, 1750-1756

Opmeer EM, Kortekaas R, Tol M, Renken RJ, Demenescu LR, Woudstra S, Ter Horst GJ, Buchem MA, der Wee NJ, Veltman DJ (2015). CHANGES IN REGIONAL BRAIN ACTIVATION RELATED TO DEPRESSIVE STATE: A 2 ‐ YEAR LONGITUDINAL FUNCTIONAL MRI STUDY. *Depression and anxiety* **33**, 35-44

Penninx BWJH, Beekman ATF, Smit JH, Zitman FG, Nolen WA, Spinhoven P, Cuijpers P, De Jong PJ, Van Marwijk HWJ, Assendelft WJJ (2008). The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *International Journal of Methods in Psychiatric Research* **17**, 121-140

Peselow ED, Corwin J, Fieve RR, Rotrosen J, Cooper TB (1991). Disappearance of memory deficits in outpatient depressives responding to imipramine. *Journal of affective disorders* **21**, 173-183

Ramel W, Goldin PR, Eyler LT, Brown GG, Gotlib IH, McQuaid JR (2007). Amygdala reactivity and mood-congruent memory in individuals at risk for depressive relapse. *Biological psychiatry* **61**, 231-239

Redlich R, Bürger C, Dohm K, Grotegerd D, Opel N, Zaremba D, Dannlowski U (2017). Effects of electroconvulsive therapy on amygdala function in major depression – a longitudinal functional magnetic resonance imaging study. *Psychological Medicine* **47**, 2166-2176. doi:10.1017/S0033291717000605

Ritchey M, Dolcos F, Eddington KM, Strauman TJ, Cabeza R (2011). Neural correlates of emotional processing in depression: Changes with cognitive behavioral therapy and predictors of treatment response. *Journal of psychiatric research* **45**, 577-587

Sapolsky RM (2000). Glucocorticoids and hippocampal atrophy in neuropsychiatric disorders. *Archives of General Psychiatry* **57**, 925-935

Schmaal L, Veltman DJ, van Erp TG, Sämann P, Frodl T, Jahanshad N, Loehrer E, Tiemeier H, Hofman A, Niessen W (2015). Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Molecular psychiatry*

Sheline YI, Barch DM, Donnelly JM, Ollinger JM, Snyder AZ, Mintun MA (2001). Increased amygdala response to masked emotional faces in depressed subjects resolves with antidepressant treatment: an fMRI study. *Biological psychiatry* **50**, 651-658

Sheline YI, Sanghavi M, Mintun MA, Gado MH (1999). Depression duration but not age predicts hippocampal volume loss in medically healthy women with recurrent major depression. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **19**, 5034-5043

Sternberg DE, Jarvik ME (1976). Memory functions in depression: Improvement with antidepressant medication. *Archives of General Psychiatry* **33**, 219-224

Tomioka H, Yamagata B, Kawasaki S, Pu S, Iwanami A, Hirano J, Nakagome K, Mimura M (2015). A longitudinal functional neuroimaging study in medication-naive depression after antidepressant treatment. *PloS one* **10**, e0120828

Treadway MT, Waskom ML, Dillon DG, Holmes AJ, Park MTM, Chakravarty MM, Dutra SJ, Polli FE, Iosifescu DV, Fava M, Gabrieli JDE, Pizzagalli DA (2015). Illness Progression, Recent Stress, and Morphometry of Hippocampal Subfields and Medial Prefrontal Cortex in Major Depression. *Biological psychiatry* **77**, 285-294

Tulving E (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne* **26**, 1

Turk-Browne NB, Yi D, Chun MM (2006). Linking Implicit and Explicit Memory: Common Encoding Factors and Shared Representations. *Neuron* **49**, 917-927

Usami M, Iwadare Y, Kodaira M, Watanabe K, Saito K (2014). Near infrared spectroscopy study of the frontopolar hemodynamic response and depressive mood in children with major depressive disorder: a pilot study. *PloS one* **9**, e86290

van Tol M, Demenescu LR, van der Wee, Nic JA, Kortekaas R, Marjan MA N, Boer JD, Renken RJ, van Buchem MA, Zitman FG, Aleman A (2012). Functional magnetic resonance imaging correlates of emotional word encoding and recognition in depression and anxiety disorders. *Biological psychiatry* **71**, 593-602

van Tol MJ, van der Wee NJA, Demenescu LR, Nielen MMA, Aleman A, Renken R, van Buchem MA, Zitman FG, Veltman DJ (2011). Functional MRI correlates of visuospatial planning in out-patient depression and anxiety. *Acta Psychiatrica Scandinavica* **124**, 273-284

Van Wingen GA, Van Eijndhoven P, Cremers HR, Tendolkar I, Verkes RJ, Buitelaar JK, Fernández G (2010). Neural state and trait bases of mood-incongruent memory formation and retrieval in first-episode major depression. *Journal of psychiatric research* **44**, 527-534

Veer IM, Beckmann CF, van Tol MJ, Ferrarini L, Milles J, Veltman DJ, Aleman A, van Buchem MA, van der Wee NJ, Rombouts SA (2010). Whole brain resting-state analysis reveals decreased functional connectivity in major depression. *Frontiers in systems neuroscience* **4**, 10.3389/fnsys.2010.00041. eCollection 2010

Victor TA, Furey ML, Fromm SJ, Öhman A, Drevets WC (2010). Relationship between amygdala responses to masked faces and mood state and treatment in major depressive disorder. *Archives of General Psychiatry* **67**, 1128-1138

Warshaw MG, Dyck I, Allsworth J, Stout RL, Keller MB (2001). Maintaining reliability in a long-term psychiatric study: an ongoing inter-rater reliability monitoring program using the Longitudinal Interval Follow-Up Evaluation. *J Psychiatr Res* **35**, 297–305

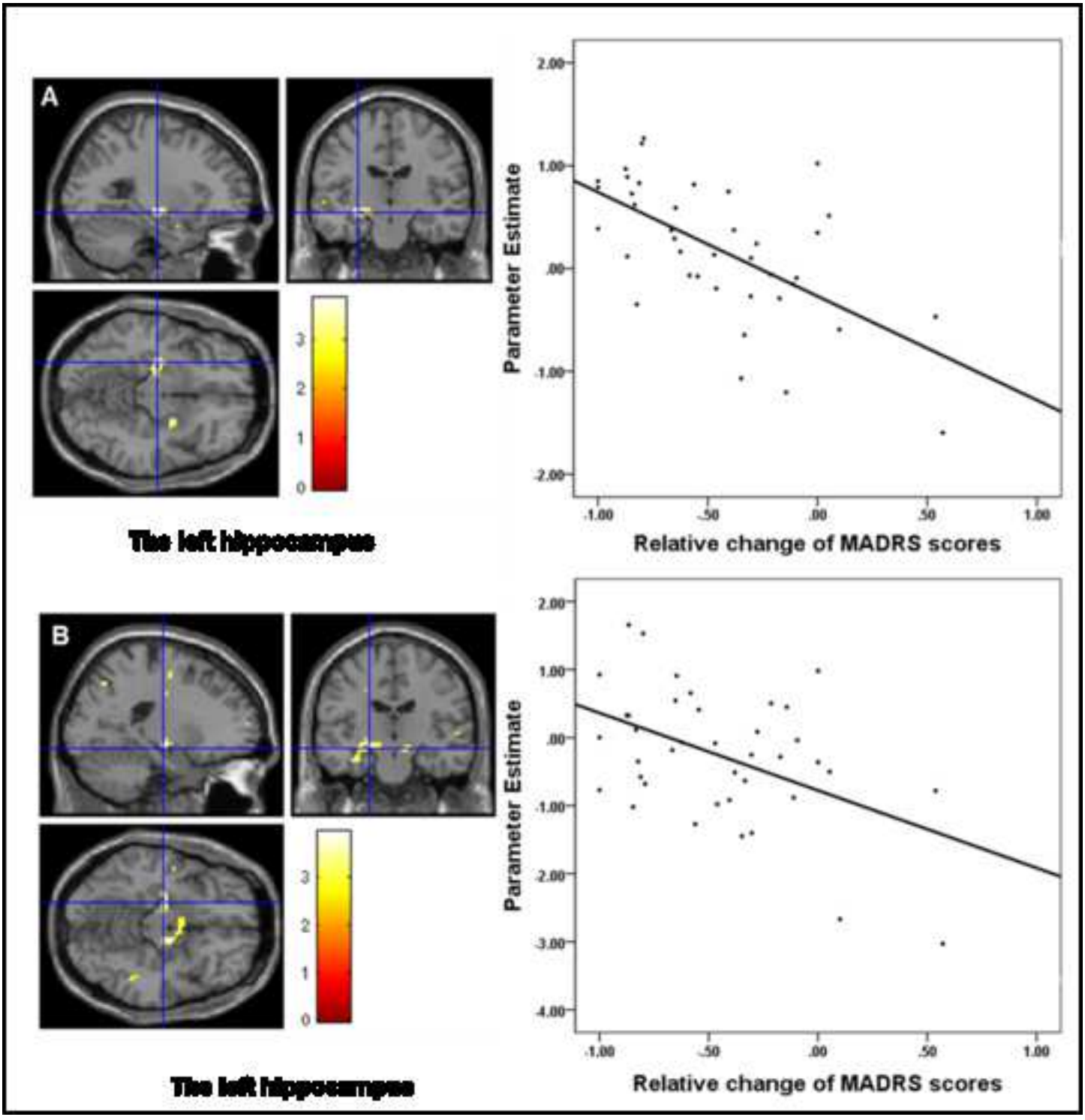
Wise T, Cleare AJ, Herane A, Young AH, Arnone D (2014). Diagnostic and therapeutic utility of neuroimaging in depression: an overview. *Neuropsychiatric disease and treatment* **10**, 1509

Young KD, Bellgowan PS, Bodurka J, Drevets WC (2015). Functional Neuroimaging Correlates of Autobiographical Memory Deficits in Subjects at Risk for Depression. *Brain sciences* **5**, 144-164

Zaremba D, Dohm K, Redlich R (2018). Association of Brain Cortical Changes With Relapse in Patients With Major Depressive Disorder. *JAMA Psychiatry* **75**, 484-492. doi:10.1001/jamapsychiatry.2018.0123

Zimmerman M, Posternak MA, Chelminski I (2004). Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. *Journal of psychiatric research* **38**, 577-582

Figure 1



Longitudinal brain changes in MDD during emotional encoding: effects of presence and persistence of symptomatology

Supplement

Supplementary methods

fMRI data acquisition and processing. A SENSE-6 channel head coil was used at S1 in Amsterdam. A SENSE-8 channel head coil was used in Groningen and Leiden at both S1 and S2 and in Amsterdam at S2. In Groningen, echo planar imaging (EPI) volumes of 39 slices were acquired using a T2*- weighted gradient echo sequence (TR=2300 ms, TE=28 ms, matrix size: 64 × 64, plane resolution: 3 × 3 mm, slice thickness: 3 mm) at S1 and the EPI slice setting was changed into 35 slices at S2. In Leiden and Amsterdam, 35 axial slices were obtained using a T2*- weighted gradient echo sequence (TR=2300 ms, TE = 30 ms, matrix size: 96 × 96, plane resolution: 2.29 × 2.29 mm, slice thickness: 3 mm) at S1 and S2. Transversal slices were acquired parallel to the anterior commissure-posterior commissure plane (no gap) in interleaved order.

In addition, a high-resolution anatomical MRI was obtained with a sagittal 3D gradient-echo T1-weighted sequence for each participant (TR=9 ms, TE=3.5 ms, matrix size: 256 × 256, voxel size: 1 × 1 × 1 mm, 170 slices).

Before preprocessing, functional images were reoriented manually to the anterior-posterior commissure plane. Preprocessing consisted of slice timing, spatial realignment and co-registration of the anatomical image to the EPI image, spatial normalizing of the image to the standard Montreal Neurological Institute (MNI) space, reslicing to a 3 × 3 × 3 mm voxel size and spatial smoothing with an 8 mm full-width at half maximum Gaussian kernel. To remove low frequency noise, a high-pass filter with a cut-off of 128 s was applied to the fMRI time-series.

For the first-level analyses, two first-level models were set up for each participant, one for S1 and one for S2. To minimize the effect of motion, the absolute scan-to-scan difference for rotational and translational displacement after realignment was computed, and scans in which the displacement was larger than 0.9 mm compared to the previous scan were censored by modeling these as individual regressors [40]. Because we were interested in the valence effects and to be consistent with our previous reports [4,10], we defined the following contrasts for each model: [*successfully encoded positive words* > *successfully encoded neutral words; positive>neutral encoding*] and [*successfully encoded negative words* > *successfully encoded neutral words; negative>neutral encoding*]. The difference between the two scan sessions was calculated for each contrast by subtracting the contrast image of the first scan from the second scan (S2-S1) for every participant using the ImCalc-option implemented in SPM8. Consequently, positive activation indicates an increase of activation from S1 to S2 and negative activation a decrease of activation from S1 to S2.

Post hoc group comparison. Because effects were observed during both positive and negative vs. neutral encoding, we decided to *post-hoc* evaluate the dependency of observed effects on changes in neutral or emotional word encoding. To this end, we built a full-factorial model with S2-S1 contrast maps as dependent variables, in which valence (positive/neutral/negative>baseline) was entered as within-subject factor (with three levels) and symptom change as interacting factor with valence. No correlation was found between brain activation during neutral>baseline and symptomatic change (Supplementary Figure S2), however neither for negative > baseline and positive > baseline. It appeared that the change of brain activation depended on the difference between emotional and neutral encoding, as contrasting change in relation to positive and neutral encoding (>baseline), and negative and neutral (> baseline) encoding again showed effects in the left

hippocampus ([positive<neutral, MNI coordinates [x=-27, y=-16, z=-11], Z=4.00, p_{FWE} =.010; negative<neutral, MNI coordinates [x=-24, y=-16, z=-11], Z=3.63, p_{FWE}=.031).

In addition, to explore the change in HC group over time and to illustrate whether changes in patients reflect normalization, we built a group (HC, HI, LI) x valence (positive>neutral, negative>neutral) x time (S1, S2) model. Activation estimates from the left hippocampus/amygdala cluster from main correlational analysis for both time points for all participants were extracted from this model.

Supplementary results

Demographic characteristics. Within patients, symptom change was not associated with age (*r*=.28, *p*=.08), years of education (*r*=.14, *p*=.38) or sex (*t*=.39, *p*=.70). Moreover, symptom change was not associated with medication use (F(3, 39)=.89, *p*=.46) or psychotherapy use (F(3,39)=.61, *p*=.61) at S1 or S2. In addition, symptom change was not related to anxiety severity at S1 (BAI-score S1; *r*=-.03, *p*=.85) and depressive load in the five years before S1 (*r*=.20, *p*=.22), but trend-wise related to depression duration between S1 and S2 (*r*=.29, *p*=.07). Symptom change of depression was correlated to change in anxiety severity (BAI-scores) (*r*=.46, *p*=.003) and depression severity at baseline (MADRS-S1; *r*=-.34, *p*=.034).

fMRI results:

Post hoc group comparison: group (HC, HI, LI) x valence (positive>neutral, negative>neutral) x time (S1, S2) . Plotting of these estimates indicated no change in our main cluster in HC, and further indicated that changes followed a trend of normalization during positive but not negative word encoding (Supplementary Figure 3A & 3B). No interaction of time by valence or main effect of valence was observed in HC. A main effect of time in HC was observed subthreshold anterior ([x=-15,y=-10,z=-17], Z=3.49, p_{FWE} =.07).

Correlations with depression duration and course. To describe depression course in between two scan sessions, remitters (n=11), non-remitters (n=18) and recurrent group (n=10) was defined based on life chart interview. Patients who satisfied the following criteria were defined as remitters: 1) more than 3 months without symptoms or symptoms without burden within 1 year after S1, 2) without recurrence in the follow-up period. Patients who have a recurrence of symptoms for at least 1 month with burden after firstly obtaining remission were classified into recurrent group. Patients who had symptoms with burden in every month following S1 during the entire follow-up period were defined as non-remitters.

To explore whether changes in brain activation correlated with depressive course in between measurements, we divided our patients into remitters (n=11), non-remitters (n=18) and recurrent group (n=10) based on the pattern of presence of symptomatology between two scan sessions. However, course trajectory was not related to activation change during both positive and negative (>neutral) word encoding.

Quality assurance analysis**. To control for possible bias or systematic effects across the scanning sites, we conducted a quality assurance analysis using MRI Quality Control tool (MRIQC) (Esteban et al 2017). From these reports, we specifically focused on the signal-to-noise ratio on source data from subjects, whom did not change scanning site (n=34). We built a repeated measures ANOVA with site (3; AMC, LUMC, UMCG) as a between-subject factor and scanning time (2; S1, S2) as a within-subject factor. We found a significant interaction between site and time ($F_{2,57}$=3.88,**

p=.03) and a significant main effect of site ($F_{2,57}$=4.80, p=.01). No main effect of time was found (Figure S4). Post-hoc analysis showed the site effect was significant at S1 but the difference was not significant at S2. The time effect was found in UMCG participants but not at the other two sites. Consequently, we performed a second test by excluding participants from UMCG (n=11). The main results did not change during positive words encoding (P_{FWE}=.032, t=4.00, Z=3.70, MNI coordinates [x=-27, y=-13, z=-11]) but became sub-threshold during negative words encoding (P_{FWE}=.120, t=3.47, Z=3.26, MNI coordinates [x=-24, y=-13, z=-11]).

Control for site effects. **To test if the results were influenced by changes of scanner sites, we repeated the analyses after excluding all participants who switched scanning site at S2 (n=5). The results showed that the correlation between changes of depressive state and brain activation change during positive emotional words encoding did not change (P_{FWE}=.028, t=3.97, Z=3.73, MNI coordinates [x=-27, y=-13, z=-11]), indicating that reported effects were not affected by changing scanning site. The correlation during negative words encoding became sub-threshold (P_{FWE}=.099, t=3.49, Z=3.32, MNI coordinates [x=-30, y=-16, z=-23]).**

References

Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE* **12**, 1-21. https://doi.org/10.1371/journal.pone.0184661

Supplementary Table S1. Use of medication between two scan sessions.

	High-improved	Low-improved	*Likelihood ratio*	*p*
Benzodiazepine(yes/no)	3/16	5/15	.07	.79
Anti-depressants	1/18	0/20	1.52	.48
SSRIs	2/17	3/17	.13	.71
SNRI	0/18	3/17	4.08	.23

Supplementary Table S2. Main effect of task of emotional memory task with an initial threshold of *p*<.005 uncorrected.

			MNI Coordinate					
Regions	k	Side	BA	x	y	z	*F*	*Z*	$p_{uncorrected}$
Main effect of task									
Inferior frontal gyrus	488	L		-51	32	16	17.40	5.05	<.001
Medial frontal gyrus	38	L		-3	29	46	7.17	3.05	<.001
Superior frontal gyrus	60	L		-21	47	40	11.19	3.98	<.001
Middle temporal gyrus	17	L		-60	-43	-2	10.12	3.76	<.001
Middle temporal gyrus		R		48	-1	-16	8.29	3.34	<.001
Temporal pole	22	L		-48	11	-23	11.26	3.99	<.001
Lingual gyrus	60	L		-12	-40	-5	9.75	3.68	<.001
Lingual gyrus	58	R		9	-34	-8	9.58	3.64	<.001
Hippocampus	23	R		18	-23	-8	8.68	3.43	<.001
Hippocampus	7	L		-18	-10	-11	7.67	3.18	<.001

Supplementary Table S3. Main effect of valence and symptom change in main analysis

					MNI Coordinate					
Regions	**k[a]**	**k[b]**	**Side**	**BA**	**x**	**y**	**z**	***T***	***Z***	p_{FWE_SVC}
Main effect of valence										
Inferior frontal gyrus	**143**	**-**	**L**	**45**	**-51**	**32**	**19**	**28.62**	**4.74**	**.027***
Main effect of symptom change										
Hippocampus/amygdala	**17**	**5**	**L**	**20**	**-24**	**-13**	**-11**	**17.79**	**3.80**	**.026***

[a.] Cluster size in whole-brain analysis; [b.] Cluster size after small volume correction.

*** Significant at p<.05 FWE corrected, voxel-level after small volume correction (SVC).**

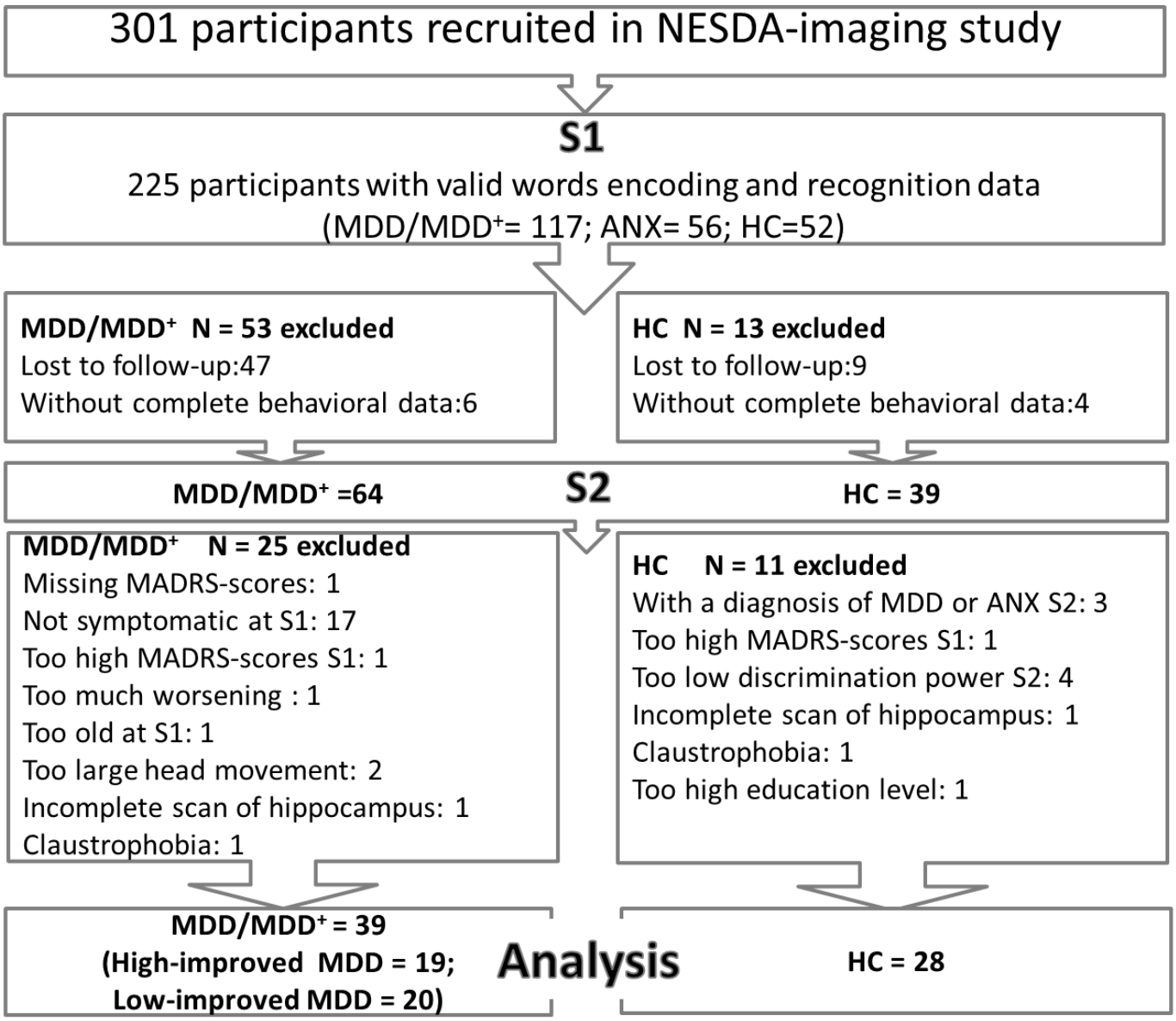
Supplementary Table S4. Effects from group × time × valence model

				MNI Coordinate					
Regions	**k**	**Side**	**BA**	**x**	**y**	**z**	***F***	***Z***	$P_{uncorrected}$
Interaction of group x time									
Inferior frontal gyrus	**11**	**R**	**44**	**57**	**17**	**31**	**10.10**	**3.84**	**<.001**
Hippocampus/amygdala	**19**	**R**	**35**	**18**	**-10**	**-17**	**9.46**	**3.70**	**<.001**

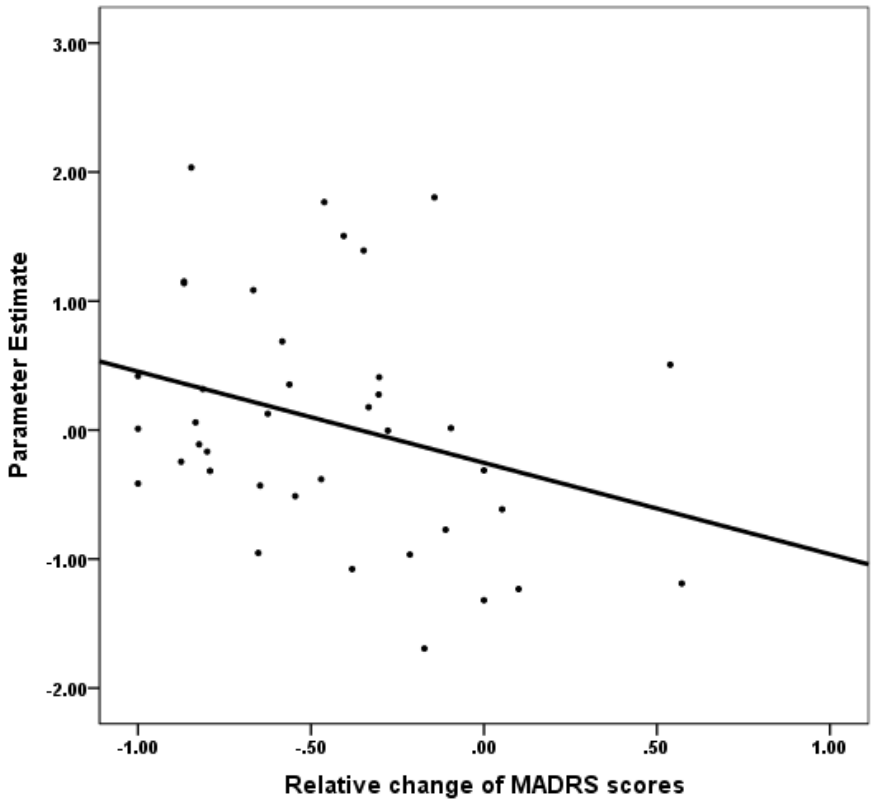
Interaction of group x time x valence									
Inferior frontal gyrus	**7**	**R**	**45**	**56**	**26**	**25**	**8.31**	**3.41**	**<.001**
Lingual gyrus	**6**	**R**	**18**	**12**	**-49**	**4**	**8.29**	**3.41**	**<.001**

Supplementary Figure S1. Flow chart of recruitment **and sample selection** of participants. **At baseline measurement (S1), 117 MDD patients and 52 healthy controls had valid scans during emotional words encoding task. From 53 of these patients and 13 of healthy controls the second measurement data (S2) could not be included because of loss to follow-up or missing behavioral data. Moreover, at S1 25 patients were not depressed at time of scanning (MADRS ≤10), one patients with too high MADRS score and 11 healthy controls were excluded to obtain a good match with the included patients. One patient was excluded based on the relative change of depressive symptom exceeded 3 standard deviation.** In total, 21 symptom-improved patients, 19 non-improved patients and 29 healthy controls were included in the final analysis.

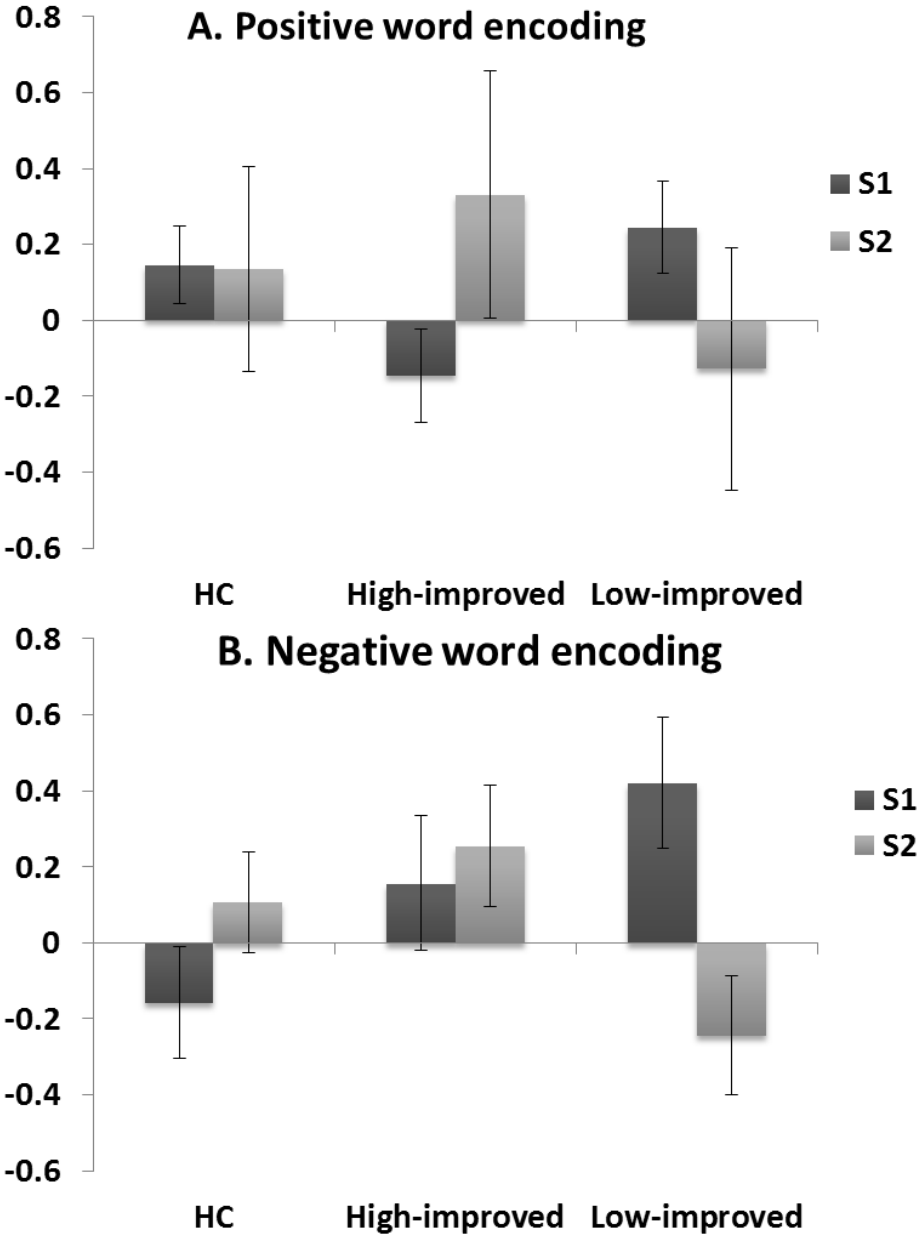
MDD: major depressive disorder; MDD^{+}: depression combined with an additional diagnosis of social anxiety disorder, panic disorder and/or generalized anxiety disorder; ANX, anxiety; S1, baseline measurement; S2, second measurement; MADRS, Montgomery–Åsberg Depression Rating Scale; HC, healthy control; S-R: symptomatic-remitted MDD patients; S-S: symptomatic-symptomatic MDD patients.



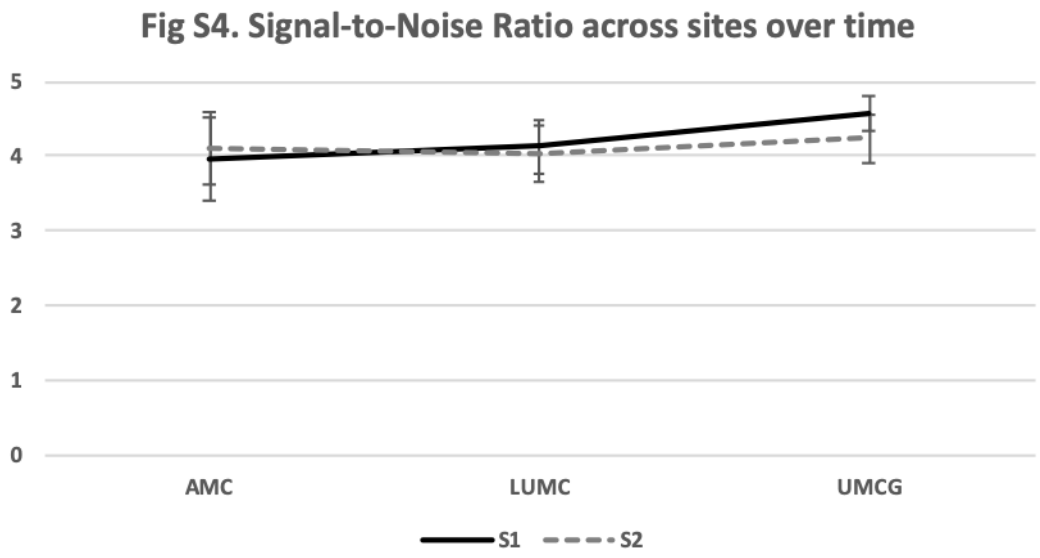
Supplementary Figure S2. No correlation of change in the hippocampus/amygdala and symptom change in MDD was found between brain activation during neutral>baseline and symptomatic change.



Supplementary Figure S3. Activation of the left hippocampal/amygdala extracted from correlation analysis during A) positive (peak MNI coordinate: x=-28, y=-13, z=-11), and B) negative word encoding at S1 and S2 (peak MNI coordinate: x=-23, y=-13, z=-11).



Supplementary Figure S4. No main effect of time was found in signal-to-noise ratio over time.



Authors' responses:

We are grateful for the reviewers' time and their insightful comments on our manuscript on longitudinal changes of brain activation in depression.
We carefully revised the manuscript and addressed each comment raised by the reviewers. Below you can find an overview of the comments, with our point by point responses. Changes in the manuscript are in **bold** font.

Reviewer #1: This longitudinal study aimed at assessing the impact of changes in depressive severity on hippocampal and amygdalar activation during memory encoding. The paper is overall very well written. However, I think it would benefit from a few clarifications:

1.1 The sentence on page 4 "no specific treatment was delivered but treatment delivery was monitored" seems contradictory. I suggest rephrasing.

Reply to 1.1: We thank the reviewer for his/her thoughtful comments. We agree that the sentence could lead to confusion. We meant to underline that treatment was not offered as part of the study but the treatment that patients received anyhow was reconstructed as part of this naturalistic longitudinal cohort study. We rephrased the sentence in our introduction, which now reads (cf. Introduction, page4, line77-79):
"In this interval, **no specific treatment was delivered as part of the study protocol.** ~~**no specific treatment was delivered, but**~~ **Given the naturalistic design of our study, participants could receive treatment as usual, which was reconstructed retrospectively based on self-reports at the two-year follow-up interview.**"

1.2 Can the authors give more details with respect to comorbidity status (e.g. how many people had comorbid generalized anxiety, social anxiety, panic disorder at S1 and S2? What was the reason for allowing comorbidity? Did the comorbidity status, change over time, from S1 to S2? Were symptoms measured for the additional comorbid conditions and considered in the analysis?

Reply to 1.2: Comorbidity was allowed because NESDA specifically aimed to gain insight into the long-term course of depression and anxiety disorders, which are highly comorbid and may share neural underpinnings. Previously we reported on cross-sectional commonalities in the neural correlates of emotional word processing amongst patients with depression and anxiety disorders in this sample (e.g. van Tol *et al.* 2012, *Biol Psychiatry*). Only comorbidity of generalized anxiety disorder, social anxiety, and/or panic disorder was allowed in the NESDA Neuroimaging study. In the current analyses, we were interested in the long-term effects of depression persistence on the neural correlates of emotional word processing for later recognition. We controlled in our analyses for the presence of comorbid anxiety status by measuring anxiety state using BAI scores and including relative change of BAI scores in the analyses. We now listed the details of comorbidity status at follow-up and their

change over time in the table below:

		HC	High-improved MDD	Low-improved MDD	F	t	χ^2	*Likelihood ratio*	*p*
N		28	19	20	-	-	-	-	
Comorbidity_S1									
Comorbid SAD(yes/no)	N	-	6/13	9/11	-	-	.74	-	.51
Comorbid PD(yes/no)	N	-	6/13	6/14	-	-	.01	-	.92
Comorbid GAD(yes/no)	N	-	7/12	10/10	-	-	67	-	.52
Comorbidity at follow-up									
Comorbid SAD(yes/no)	**N**	**-**	**2/17**	**6/14**	**-**	**-**	**-**	**2.36**	**.13**
Comorbid PD(yes/no)	**N**	**-**	**2/17**	**6/14**	**-**	**-**	**-**	**2.36**	**.13**
Comorbid GAD(yes/no)	**N**	**-**	**0/19**	**8/12**	**-**	**-**	**-**	**12.66**	**<.01**

[1.] HC differed from both patient groups, while the two patient groups did not differ; [2.] All groups differed from each other; [3.] Infrequent use; [4.] Two patients used benzodiazepine frequently. * significant at $p<.05$

HC: healthy control; S-R: symptom-remitted MDD patients; S-S: symptomatic-symptomatic MDD patients; SAD: social anxiety disorder; PD: panic disorder; GAD: generalized anxiety disorder.

We observed that comorbidity of MDD with SAD and PD did not differ between low and high improved MDD patients at S1, though GAD at time of follow up was only present in the low improved group. We therefore added to our discussion a comment that our results could be potentially affected by comorbidity status, in particular GAD.

Adjustment in the manuscript:

(cf. Discussion, page11, line330) "**Furthermore, although comorbidity of SAD and PD was similar in low and high improved MDD groups, GAD was more frequent in low-improved MDD patients, which may have affected our results."**

1.3 The reason for including some medication use and drawing the line for certain dosage is not very well explained. What was the reason for accepting SSRIs and benzodiazepines only? How rigid was this selection and why? Even though in small numbers, the supplementary material shows the use of SNRIs, as well as "other antidepressants". What was the rationale for choosing the cut-off point of 20 mg of benzodiazepine up to 3 times/week?

Reply to 1.3: We chose not to aim for medication-free participants only, to ensure that the patients participating in the MRI sub-study were a representative sample of the overall NESDA cohort. However, psychotropic medication was limited to stable use of SSRI's and/or infrequent use of benzodiazepines, equivalent to 20mg oxazepam (the most widely prescribed anxiolytic drug in the Netherlands) three times a week or less. This was done to reduce variance associated with use of e.g. tricyclic antidepressants, MAO-inhibitors or atypical antipsychotics (i.e., SSRI's only) and because frequent/chronic benzodiazepine use is a likely confound when presenting 'emotional' stimuli (words or faces). Therefore, participants were requested to abstain from

benzodiazepine use 48 hours prior to scanning. The reviewer correctly notes that at S2, three patients used SNRIs: this was due to the observational nature of the NESDA study in which 'treatment as usual' was permitted, without study-specific treatment guidelines (cf. our reply to 1.1).

1.4 I am unclear how the percentage of time spent with depression and presence of depressive symptoms per month between S1 and S2 was measured. What exactly was recorded in the life chart review, how often, by whom etc. Was it all assessed at S2? If yes, did the authors consider a recall bias? This was not mentioned in the limitations.

Reply to 1.4: The Life Chart Method (LCM) was administered at S2 by trained professionals to retrospectively assess presence and severity of symptoms. The LCM starts with exploring the presence of life events in a certain period to refresh memory. Subsequently, presence and severity of depressive symptoms (no burden, small burden etc.) is assessed (Lyketsos*, et al* 1994). For each participant, the total number of months with depressive symptoms with at least small burden within the follow-up period was computed. The life chart interviews were performed at both the baseline and two-year follow-up interview session. The baseline LCM was used to determine the presence of symptoms in the five years (per year) prior to baseline. The follow-up LCM administered at the two-year follow-up interview was used to calculate percentage of time spent with depression during this two-year-course. The validity and reliability of its methodology have been shown to be good among patients (Warshaw*, et al* 2001).

Nevertheless, we agree that retrospective reports could be subject to recall bias. We now acknowledged this possible limitation in the Discussion section.

Adjustment in the manuscript:

(cf. Methods and Materials, page6, line155) "Percentage of months experiencing depressive symptoms **relative to the overall follow-up period** was calculated per patient as time spent with depression (Ai *et al* 2015)."

(cf. Discussion section, page11, line328-330) "**Next, the retrospective life chart method used to measure persistence of depressive symptoms might have been subject to patients' mood state, though the reliability and validity have been estimated to be relatively high (Warshaw, *et al* 2001).**"

1.5 After the sensitivity analysis was done, it seems that 6 more patients were excluded. What was the final number of patients in each of the groups

Reply to 1.5: In the main analyses, we performed all analyses on all depressed patients, which included patients who were equally or more depressed at S2 than at S1. As a reliability check, we did the sensitivity analyses for both behavioral and fMRI

data only within patients who showed symptomatic improvement. Consequently, the sample size for the continuous sensitivity analysis changed from 39 to 33. The sensitivity analyses were thus only performed for the continuous analyses, and not for the group analyses.

Adjustment in the manuscript:
(cf. Method, page 7, line194-195) “A sensitivity analysis was planned to test whether associations **would hold in the analysis including only** patients with symptomatic improvement **(n=33).**”

1.6 In the discussion the authors also mention that "no other regions were found to change as a function of symptomatic improvement". Did they actually explore this as well? The introduction states that they mostly focused on activation of the amygdala and hippocampus.

Reply to 1.6: Indeed, our regions of interests were the amygdala and hippocampal areas and in addition we explored effects using whole-brain analyses. This means that due to our specific hypotheses, we *a priori* considered effects occurring in these regions significant at $p<.05$ FWE corrected for the extent of a spatial mask encompassing all voxels covering the bilateral hippocampi and amygdalae. Nevertheless, we explored whether activation in other voxels outside these regions were additionally associated with severity and time spent with depressive symptoms. For effects occurring outside the amygdala and hippocampus we set the threshold $p<.05$, FWE whole-brain corrected, but no areas survived. We now make this more explicit in our introduction and methods.

Adjustments in the manuscript
(cf. Introduction, page4, line81-82): “~~**We further aimed to**~~ **Furthermore, we aimed to** explore whether activation in ~~**other brain**~~ regions **other than amygdala and hippocampus** ~~(such as ventromedial prefrontal cortex, anterior cingulate cortex and frontal pole)~~ related to longitudinal treatment responses was associated with severity and time spent with depressive symptoms.”

(cf. Methods, page8, line213): “**We also examined the effects in other brain regions than ROIs, which** ~~**Effects occurring outside the amygdala and hippocampus**~~ had to meet p<.05, FWE whole-brain corrected to be considered significant.”

1.7 I am not convinced that showing how emotional memory formation is sensitive to changes in depression severity, can justify the conclusion that hippocampal and amygdalar brain activation is not subject to functional scarring. Would one single study, with all the listed limitations, be able to support such a claim?

Reply to 1.7: These conclusions were based on results from two main analyses: 1) We found that brain activation changes in the hippocampus and amygdala were associated with changes in depression severity, which suggested *state-dependency* of these functional abnormalities in depression; 2) We found no relation between time spent with depression and activation changes in the hippocampus and amygdala, which we interpreted as an indication that functional abnormalities are not associated with longer depressive duration and therefore are not subject to 'functional scarring'.

We agree, however, that these conclusions are based on a single study (with limitations) and clearly in need of replication. We therefore have toned down this conclusion, which now reads (cf. Conclusion, page 12, line339): "Taken together, **our results suggest** that hippocampal activation is a state-dependent characteristic that **is not related to persistence of depression. This may indicate that functional activation patterns in depression are** not subject to functional 'scarring', **a hypothesis that deserves further investigation**."

Adjustments in the manuscript

(cf. Abstract, page2, line33-36): "*Conclusion:* Using a longitudinal within-subjects design we showed that hippocampal-amygdalar activation during emotional memory formation is related to depressive symptom severity **but not persistence (i.e. time spent with depression or 'load')** ~~was not associated with activation changes over time~~, suggesting **functional activation patterns in depression** ~~that hippocampal- and amygdalar brain activation~~ **are** not subject to functional 'scarring' **although this hypothesis awaits future replication**."

Reviewer #2: In the study "Longitudinal brain changes in MDD during emotional encoding", the authors investigated if changes in brain activation during emotional encoding was associated with symptomatic improvement and persistence of depressive symptoms. This research question of great importance because it remains unclear from previous studies if functional changes in patients with MDD are state-dependent or not. Therefore, the authors conducted a 2-year follow-up longitudinal study with MRI measurements at both time-points. Furthermore, pharmacological treatment and psychotherapy during the follow-up interval were monitored and included in statistical analyses, which is a major strength of the study. Apart from the important research question and the selection of a well-suited study design, the manuscript is also well written.

Although the statistical models are well suited to answer the research questions, the results (of both main and supplementary analyses) should be reported conclusively with a systematic presentation of all main effects and interactions. Furthermore, the authors should comment on the signal stability over time. Because data was acquired at different scanners and some participants even switched scanning site from baseline to follow-up, I am concerned how reliable contrast maps between follow-up and baseline fMRI are.

2.1 Introduction
Second paragraph: the authors may consider citing other interventions than psychotropic medication such as electroconvulsive therapy when discussing the effects of treatment on amygdalar and hippocampal activation (e.g. Redlich et al., 2017).

Reply to 2.1: We thank the reviewer for taking time to thoroughly review our manuscript and pointing out important issues. We agree that effects of ECT on MTL activation deserve mentioning and have adapted our Introduction as follows (Introduction, page3, line55):

“Findings have been inconclusive with reports of decreased **(Fu *et al* 2004, Sheline *et al* 2001, Redlich *et al* 2017**), increased (Goldapple *et al* 2004, Neumeister *et al* 2006, Ritchey *et al* 2011, Victor *et al* 2010), or unchanged (Fu *et al* 2015, Opmeer *et al* 2015) activation following successful short-term pharmacological treatment (Fu *et al* 2004, Sheline *et al* 2001, Victor *et al* 2010, Fu *et al* 2015), **electroconvulsive therapy (Redlich *et al* 2017)**, cognitive behavioral treatment (Fu *et al* 2008, Goldapple *et al* 2004, Ritchey *et al* 2011), or naturalistic remission (Opmeer *et al* 2015).”

2.2 Methods

2.2.1 The information about how experimental conditions were contrasted for fMRI analyses should be stated more explicitly in the main body of the manuscript, not only in the supplementary materials. The wording "positive > neutral" and "negative > neutral" is misleading because the authors have investigated only successfully encoded stimuli based on the recognition task. Please include this information in the methods section.

Reply to 2.2.1: Indeed, we only investigated successfully encoded stimuli during the memory encoding phase. We adapted the relevant wording in our manuscript accordingly.

Adjustments in the manuscript:

(cf. Methods session, page7, line179) “To test for the association between symptom change and change of brain activation during positive and negative encoding over the two-year interval, scan moments, S2-S1 contrast maps were entered as dependent variables in a full-factorial model, with valence (**successfully encoded positive words>successfully encoded neutral words** ~~**positive>neutral encoding**~~ [S2-S1], **successfully encoded negative words>successfully encoded neutral words** ~~**negative>neutral encoding**~~ [S2-S1]) as a factor and symptom change ([MADRS S2 – MADRS S1]/MADRS S1) as interacting factor with valence.”

(cf. Methods session, page8, line198) “We built a full factorial model with valence as factor (2; **successfully encoded positive words>successfully encoded neutral words and successfully encoded negative words>successfully encoded neutral words** ~~**positive>neutral and negative>neutral**~~) and time spent with depressive symptoms as an interacting covariate with valence.”

(cf. Results session, page9, line238) “Adding time spent with depressive symptoms in the interval between S1 and S2 as covariate did not change the results (Z=3.85, p_{FWE}=.019 for **successfully encoded positive words>successfully encoded neutral words** ~~**positive>neutral encoding**~~ (pos); Z=3.95, p_{FWE}=.014 for **successfully encoded negative words>successfully encoded neutral words** ~~**negative>neutral encoding**~~ (neg)).”

(cf. Results session, page9, line252) “No correlation between percentage of time with depressive symptoms and changes in brain activation was observed across all MDD patients during **successful encoding of positive and negative words.** ~~positive>neutral and negative > neutral encoding~~”

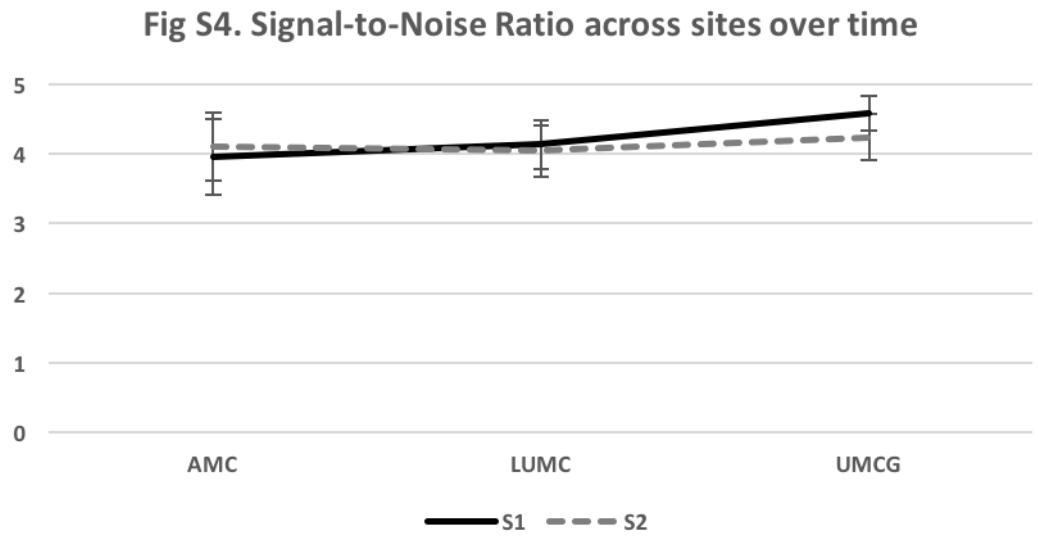
2.2.2 Although the authors included covariates of no interest to account for participants who switched scanning sites during the study protocol, I am still concerned about the data quality and reliability of follow-up measurements from different scanners. The authors mention "minor" variations in sequence and coil but do not comment on assurance of signal stability and reliability. Has there been any protocol for MRI quality assurance?
Because the authors calculated contrast maps to indicate change in brain activation, it would be important to comment on scanner/signal stability over time and between scanning sites. I also suggest to exclude all participants who changed scanning site at follow-up from the analyses to see if the reported effects are independent of scanning site (if so, these analyses may be shifted to the supplementary materials).

Reply to 2.2.2: We agree with the reviewer that signal stability and reliability is very important to check in longitudinal studies. We took a number of measures to investigate how scanner variability and signal stability over time could have affected our results.

First, to explore how changing scanners affected the results, we excluded all participants who changed scanning site (n= 5) and repeated the analyses. The results showed that the correlation between brain activation change and positive emotional words encoding did not change (P_{FWE}=.028, t=3.97, Z=3.73, MNI coordinates [x=-27, y=-13, z=-11]), indicating that reported effects were not affected by change in scanning site. The correlation during negative words encoding however became sub-threshold (P_{FWE}=.099, t=3.49, Z=3.32, MNI coordinates [x=-30, y=-16, z=-23]).

Furthermore, to check the signal stability and reliability, we did a MRIQC analysis with data of subjects whom did not change scanning site over time (n=34). After we calculated the signal-to-noise ratio (SNR) on the raw data, we built a repeated measures ANOVA with site (3; AMC, LUMC, UMCG) as a between-subject factor and scanning time (2; S1, S2) as a within-subject factor. There was a significant interaction between site and time ($F_{2,57}$=3.88, p=.03) and a significant main effect of site ($F_{2,57}$=4.80, p=.01). No main effect of time was found. Post-hoc analysis showed the site effect was significant at S1 but the difference was not significant at S2. Specifically, SNR at UMCG showed significant differences with SNRs at AMC (p=.011) and LUMC (p=.044). Also, post-hoc analysis showed a time effect for SNR at UMCG scans only. These results led us to perform the following sensitivity analyses in our main analyses :1) we controlled for site by adding four dummy variables (i.e., both time scanned in AMC; changed from AMC to LUMC; changed from LUMC to AMC; both time scanned in UMCG; both time scanned in LUMC) and as shown in our original submission, this did not affect our main results, 2) Next we excluded all subjects from the UMCG (n=11) to explore if main effects would be affected by this. The main results did not change during positive words encoding (P_{FWE}=.032, t=4.00, Z=3.70, MNI coordinates [x=-27, y=-13, z=-11]) but became sub-threshold during negative words encoding (P_{FWE}=.120, t=3.47, Z=3.26, MNI coordinates [x=-24, y=-13, z=-11]).

In summary, the MRIQC analyses therefore showed that in Leiden and Amsterdam, the signal was fairly stable, though in Groningen SNR slightly dropped in the 11 participants included at that site (Fig S4). However, excluding these patients from the main analysis still indicated an effect during positive word encoding and a slightly weaker effect during negative encoding, which could also relate to loss of power. We therefore think that our observations, especially those observed during positive encoding, were not primarily driven by site-specific changes in signal over time. We added these analyses to supplementary results and discussed the scanner change and signal stability as limitations.



Adjustments in the manuscript:

(cf. Discussion, page11, line324) "**Fifth, although the site effect was controlled by adding it as a covariate, it might still have confounding effect on our results. Quality assurance analysis and exploration by excluding patients that switched scanners between measurements (supplementary results) revealed similar results. These indicate that our observed effects, especially those observed during positive encoding, were not primarily driven by site-specific changes in signal over time.**"

(cf. Supplementary results) "***Quality assurance analysis*. To control for possible bias or systematic effects across the scanning sites, we conducted a quality assurance analysis using MRI Quality Control tool (MRIQC) (Esteban et al 2017). From these reports, we specifically focused on the signal-to-noise ratio on source data from subjects, whom did not change scanning site (n=34). We built a repeated measures ANOVA with site (3; AMC, LUMC, UMCG) as a between-subject factor and scanning time (2; S1, S2) as a within-subject factor. We found a significant interaction between site and time ($F_{2,57}$=3.88, p=.03) and a significant main effect of site ($F_{2,57}$=4.80, p=.01). No main effect of time was found (Figure S4). Post-hoc analysis showed the site effect was significant at S1 but the difference was not significant at S2. The time effect was found in**

UMCG participants but not at the other two sites. Consequently, we performed a second test by excluding participants from UMCG (n=11). The main results did not change during positive words encoding (P_{FWE}=.032, t=4.00, Z=3.70, MNI coordinates [x=-27, y=-13, z=-11]) but became sub-threshold during negative words encoding (P_{FWE}=.120, t=3.47, Z=3.26, MNI coordinates [x=-24, y=-13, z=-11]).

***Control for site effects*. To test if the results were influenced by changes of scanner sites, we repeated the analyses after excluding all participants who switched scanning site at S2 (n=5). The results showed that the correlation between changes of depressive state and brain activation change during positive emotional words encoding did not change (P_{FWE}=.028, t=3.97, Z=3.73, MNI coordinates [x=-27, y=-13, z=-11]), indicating that reported effects were not affected by changing scanning site. The correlation during negative words encoding became sub-threshold (P_{FWE}=.099, t=3.49, Z=3.32, MNI coordinates [x=-30, y=-16, z=-23]).**"

2.2.3 The authors report that they performed separate analyses for the positive and negative condition with regard to the effects of persistence of depressive symptoms. It seems to me that the same holds true for the effects of change of depressive state, but this information is not clear to me from the methods section. Please clarify.

Reply to 2.2.3: Indeed, the analyses on effect of change of depressive state were also conducted for positive and negative words encoding separately. We now made clarified this issue in the methods session, which now reads:
(cf. page7, line182) "**Contrast maps were built for successful encoding of positive words (vs. successful encoding of neutral words) and negative words separately.**"

2.3 Results
2.3.1 In the results, the authors report a negative correlation between symptomatic improvement and both positive and negative word encoding. If both conditions show a correlation with symptomatic improvement separately, the main effect of symptomatic improvement should be significant as well. Please report the main effect of symptomatic improvement (irrespective of condition).
Please also report the main effect of valence and the interaction of valence x symptom change.

Reply to 2.3.1: In our main analysis, we built the correlation model with valence as a factor and symptom improvement as an interacting factor with valence, which allowed us to investigate the effect of symptom change on positive and negative condition separately and may reflect the interaction between symptom change and valence.
In our revision, we built a new model in order to report the main effects of valence and

symptom change, with these two variables as two independent (non-interacting) factors. The results were listed below and added to the supplementary table (S3). There was no interaction between valence and symptom change.

					MNI Coordinate					
Regions	k[a]	k[b]	Side	BA	x	y	z	*T*	*Z*	p_{FWE_SVC}
Main effect of valence										
Inferior frontal gyrus	143	-	L	45	-51	32	19	28.62	4.74	.027*
Main effect of symptom change										
Hippocampus/amygdala	17	5	L	20	-24	-13	-11	17.79	3.80	.026*

[a.] Cluster size in whole-brain analysis; [b.] Cluster size after small volume correction.

* Significant at p<.05 FWE corrected, voxel-level after small volume correction (SVC).

2.3.2 MNI coordinates given in Figure 1 do not correspond to the cluster coordinates given in Table 2. However, the authors state that the voxels selected for Figure 1 are the peak voxels derived from the correlations of symptom change and brain activation. Which coordinates are shown in Table 2? Please clarify.

Reply to 2.3.2: After we double-checked the peak coordinates and figure images, we found that the peak voxel in the results table (Table 2) during successful encoding of positive words should be updated as [-27, -16, -11] and the peaks in the legend of Figure 1 should be [-27, -16, -11] and [-24, -13, -11]. Despite the inconsistency in the legend, the coordinates were correctly presented. We apologize for any confusion resulting from the inconsistency between the legend of Figure 1 and Table 2.

Adjustments in the manuscript:

(c.f. Figure 1. Legend) “Brain activation during emotional word encoding. A). Negative association between symptom change and hippocampal activation change during positive word encoding. (peak MNI coordinate: x=~~-28~~**-27**, y=~~-13~~**-16**, z=-11); B). Negative association between symptom change and hippocampal activation change during negative word encoding. (peak MNI coordinate: x=~~-23~~**-24**, y=-13, z=-11).”

(cf. Table 2)

Table 2. Correlation between state-change scores and brain activation changes across patients

					MNI Coordinate					
Regions	k[a]	k[b]	Side	BA	x	y	z	*T*	*Z*	p_{FWE_SVC}

~~*Positive>neutral*~~ ***successfully encoded positive words>successfully encoded neutral words****:* *negative correlation*										
Hippocampus/amygdala	35	13	L	20	~~-23~~**-27**	~~-18~~**-16**	-11	3.83	3.63	.040*
Hippocampus/amygdala	33	9	R	34	27	-4	-11	3.46	3.31	.107
~~*Negative>neutral*~~ ***successfully encoded negative words>successfully encoded neutral words:*** *negative correlation*										
Hippocampus/amygdala	50	22	L	-	-24	-13	-11	3.76	3.57	.049*
Hippocampus/amygdala	59	20	R	-	15	-7	-17	3.40	3.26	.122

2.3.3 In the supplementary materials, the authors report a group x time x valence ANCOVA for the fMRI data. Why did you choose time as a within-subject factor instead of using S2-S1 contrast maps as in your correlational analyses? Please also report the interaction of group x time (if you stay with time instead of using contrast maps) and the three-way interaction of group x time x valence to see if the three groups (HC, HI, LI) show different trajectories of brain activation in response to positive and negative words.
I would expect to find F- and p-values for main effects and all possible interactions from voxelwise analyses within the hippocampus and amygdala ROI (as for the primary analyses, defined by WFU pickatlas). Instead, the authors extracted activation estimates from clusters derived by preceding correlational analysis. However, the MNI coordinates given in Supplementary figure S3 do not correspond to the cluster coordinates in Table 2. This procedure for supplementary analyses seems like double dipping to me, please comment on this.

Reply to 2.3.3: The aim for building a group × time × valence model was to follow-up our continuous analyses to additionally examine the change in HC group over time and to illustrate whether changes in the clusters resulting from our correlation analysis in patients reflected normalization. We now added the interaction of group × time and interaction of group × time × valence in the supplementary Table.

Of note, we only plotted the effects in patients and HCs for the purpose of illustrating the normalization. Also, despite that we found state-dependency of brain activation during emotional word encoding within patients, we did not observe this from a formal group × time × valence interaction. However, investigating this was not the aim of our study, because we aimed to investigate changes over time within depressed patients.

				MNI Coordinate					
Regions	k	Side	BA	x	y	z	*F*	*Z*	$P_{uncorrected}$
Interaction of group x time									
Inferior frontal gyrus	11	R	44	57	17	31	10.10	3.84	<.001
Hippocampus/amygdala	19	R	35	18	-10	-17	9.46	3.70	<.001
Interaction of group x time x valence									
Inferior frontal gyrus	7	R	45	56	26	25	8.31	3.41	<.001
Lingual gyrus	6	R	18	12	-49	4	8.29	3.41	<.001

Adjustment in the manuscript:

(cf. discussion, page11, line316): "**And this effect was not found in a formal group × time × valence interaction. However, testing this was not the aim of our paper because we focused on changes over time within depressed patients.**"

2.4 Discussion

Fifth paragraph: When you discuss structural alterations of the hippocampus, which are associated with patients' course of illness, please consider the arbitrary selection of clinical variables to characterize patients' course of illness as one source of the heterogeneous results (e.g., McKinnon et al., 2009; Zaremba et al., 2018).

Reply to 2.4: We added the related findings in the discussion, which now reads (cf. Discussion, page11, line302):

"Previous cross-sectional and longitudinal studies suggested that hippocampal volume is negatively related to duration of illness in MDD, represented by **history of psychiatric hospitalization (Zaremba *et al* 2018)**, number of episodes (MacQueen *et al* 2003, Treadway *et al* 2015) and duration of untreated illness (Sheline *et al* 1999), though not consistently (Bremner *et al* 2000, **McKinnon *et al* 2009**). At the same time, volumetric changes in the hippocampus have been linked to symptomatic improvement following treatment (Arnone *et al* 2012a), suggesting state-dependency of hippocampal volume. In the present study, though patients differed in course trajectory of depression, changes of brain activation were not related to time spent with depression, indicating that functional longitudinal changes observed in the hippocampus are load-independent. **However, the variety in selected clinical variables of current and previous studies might explain some heterogeneity in reported results.**"

Reviewer #3: This study by Ai et al. presents longitudinal fMRI findings in a cohort of 39 MDD and 28 HC subjects scanned twice over a two-year period using an emotional word-encoding and recognition task. Their aim was to identify changes in BOLD activation to emotional word encoding in the hippocampus/amygdala that were associated with symptom change and with time spent with depressive symptoms between measurements. They found a significant inverse association between BOLD activation change in L hippocampus (but not R) and symptom change - particularly during positive encoding. There was no association between BOLD activation and time spent depressed.

There are strengths to this study including a relatively large sample of depressed patients scanned twice over a two-year period, which is not an easy feat. The use of an emotional word encoding task is not particularly novel, but does build on a substantive literature of emotional bias abnormalities in depression. However, there are also some downsides including the inclusion of MDD patients with SAD, PD, or GAD and concomitant SSRI use and the naturalistic design of the study, which precludes the examination of treatment-specific effects on the imaging markers. Nonetheless, the study has merit and provides worthwhile findings that build on the existing literature in this area.

The study overall is methodologically sound, but I have one comment on the reporting of the study sample. In the bottom paragraph of p. 5, it is unclear why only 39 MDD participants were included in the final analysis when 64 MDD patients had scan data at both time points. This is not well supported in the text. Looking at Fig S1, there are several reasons provided for exclusion of these subjects that are poorly described. For example, why were 17 subjects excluded for not being symptomatic at S1? If they were not symptomatic at S1 then why were they included in the MDD group to begin with? Similarly, why was a subject excluded for having a MADRS score that was too high at S1? Why was a subject who was "too old" at S1 scanned at all? There needs to be more specifics overall for the rationale for exclusion of such a substantial fraction of your original sample at S1.

Reply to Reviewer 3: We thank the reviewer for his/her compliments and apologize if our description of our sample selection was unclear. Overall, 301 participants in the Netherlands Study of Depression and Anxiety (NESDA) were included in the MRI sub-study. Of these, as described in the baseline study by van Tol *et al.* (2010), 225 participants had valid behavioral and imaging data on emotional words encoding and recognition task, which constituted the baseline sample at S1. For this specific study, we included MDD patients who met the following criteria: diagnosis of major depressive disorder in the past half-year according to the Composite International Diagnostic Interview (CIDI life time - version 2) as assessed during the baseline interview and a current depressive state at the day of scanning (which on average

took place 8 weeks following the interview) defined as a score larger than 10 on the Montgomery–Åsberg Depression Rating Scale (MADRS). Therefore, 17 patients were excluded because they met criteria for MDD during the past half year during the initial interview but were not in a depressive state at the moment of scanning. One participant with too high MADRS scores at S1 was excluded in order to match the participants at depression severity at baseline. One additional participant was excluded because its depression severity score increased 146% from S1 to S2, which therefore constituted an outlier of clinical worsening. Finally, we excluded one patient in low-improved group to obtain a good match on all demographic variables between patients and healthy control groups. We now updated the flow chart and its descriptions to make the selection clearer.

Adjustment in the manuscript:

(cf. supplemental figure S1, Figure legend): Flow chart of recruitment **and sample selection** of participants. **At baseline measurement (S1), 117 MDD patients and 52 healthy controls had valid scans during emotional words encoding task. From 53 of these patients and 13 of healthy controls the second measurement data (S2) could not be included because of loss to follow-up or missing behavioral data. Moreover, at S1 25 patients were not depressed at time of scanning (MADRS ≤10), one patients with too high MADRS score and 11 healthy controls were excluded to obtain a good match with the included patients. One patient was excluded based on the relative change of depressive symptom exceeded 3 standard deviation.** In total, 21 symptom-improved patients, 19 non-improved patients and 29 healthy controls were included in the final analysis.